

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
ESCUELA DE POSGRADO



**ANÁLISIS DE LA TRANSPARENCIA EN LA
ELICITACIÓN DE REQUERIMIENTOS AL COMBINAR
HISTORIAS DE USUARIO Y CASOS DE USO**

Tesis para optar el grado de Magíster en Informática con mención en
Ingeniería de Software que presenta

DENNIS STEPHEN COHN MUROY

Dirigido por

DR. JOSÉ ANTONIO POW SANG PORTILLO

Jurado

MG. CLAUDIA MARÍA DEL PILAR ZAPATA DEL RÍO

DR. JOSÉ ANTONIO POW SANG PORTILLO

MG. LUIS ALBERTO FLORES GARCÍA

San Miguel, 2016



Resumen

La elicitación de requerimientos es una de las principales tareas que debe llevarse a cabo para la correcta implementación de un desarrollo software. Su incorrecta especificación genera costos innecesarios a lo largo del proyecto e inclusive, su completo fracaso. Actualmente existen diversas técnicas que sirven como apoyo a la elicitación de requerimientos; siendo las historias de usuario y el modelo de casos de uso 2 de las más conocidas.

A la fecha existen 3 propuestas metodológicas que hacen un uso combinado de ambas técnicas: Athena, K-gileRE y NORMAP. Sin embargo, ninguna de las propuestas identificadas ha sido aplicada fuera de un entorno académico.

Dados los pocos estudios que han validado la eficacia del uso combinado de los casos de uso y las historias de usuario, se optó por medir - en efectividad, eficiencia y satisfacción - el impacto en la transparencia al combinar las historias de usuario y el modelo de casos de uso. Para ello, se llevó a cabo un experimento empírico en donde participaron 25 alumnos de octavo ciclo de la especialidad de ingeniería informática de la Pontificia Universidad Católica del Perú. La evaluación requirió que los participantes desarrollen una serie de ejercicios y respondan un par de cuestionarios enfocados en medir su percepción respecto a la facilidad, utilidad e intención de uso de la combinación de ambas técnicas.

Con los resultados recabados se puede concluir que los participantes en el experimento tuvieron menos consultas y pudieron elaborar el diseño con mayor rapidez cuando trabajaron con casos de uso e historias de usuario en forma conjunta. Asimismo, los resultados del cuestionario de percepción muestran que, tanto para el análisis como el diseño, los participantes consideran que la combinación de ambas técnicas es más fácil de usar, más útil y tendrían la intención de utilizarlas en futuros proyectos.



Abstract

Requirements elicitation is one of the main tasks that must be performed in order to guarantee the correct implementation of a software development. Its incorrect specification can cause unnecessary overdue costs for the project and, in some cases, its complete failure. There are two well-known techniques that can support the requirements elicitation process: user stories and use cases model.

Until now, there are 3 methodological proposals that combine both techniques: Athena, K-gileRE y NORMAP. However, none of them have been applied in a non-academic environment.

Due to the few studies that have evaluated the efficacy of the combined use of both techniques, it was decided to evaluate - in effectiveness, efficacy and satisfaction - the impact on the transparency of the requirements elicitation process while both techniques are applied. An empirical experiment, where 25 informatic engineering students participated, was executed. The experiment required the participants to solve some exercises and answer a couple of questionnaires which are focused on measuring the perception of easiness of use, utility perception and intention of use of both techniques in combination.

Once the results were analyzed, it is possible to conclude that the participants on the experiment had fewer doubts and it was easier for them to design the software when using the uses cases model and user stories in combination. Also, the data of the perception questionnaires shows that during the software analysis and design, the participants found that using both techniques in combination was easier and more useful, and that they intend of using them in future projects.





Dedicado a mi madre.



Agradecimientos

Como autor de esta tesis quisiera extender mi agradecimiento a las siguientes personas por el apoyo brindado durante la elaboración del presente estudio.

En primer lugar, quiero agradecer a mi madre, Elizabeth Muroy, por su constante apoyo a lo largo de mis estudios desde el colegio.

A mi asesor, Toño Pow Sang, por sus comentarios y recomendaciones que me permitieron afinar el alcance y el desarrollo del presente estudio.

A la profesora Loretta Gasco por sus enseñanzas que me sirvieron para llevar a cabo el análisis estadístico de los datos del experimento. Al profesor Luis Flores y a sus alumnos del curso de Ingeniería de Software por su disposición para desarrollar el experimento. A los profesores Andrés Melgar, Abraham Dávila y Claudia Zapata porque gracias a sus enseñanzas pude comprender la profundidad de lo que implica desarrollar un trabajo de investigación.

A mis compañeros de la Dirección de Informática Académica, Oscar y Nadia, por apoyarme con la revisión de los primeros capítulos de la tesis y los instrumentos utilizados en el experimento. Sus sugerencias me permitieron acomodar los textos para facilitar su comprensión.

Finalmente, a mis compañeros de estudios Natalí Flores y José Barturén por sus palabras de aliento a lo largo de estos 2 años y su apoyo para validar el experimento. Gracias a ambos, me animé a iniciar esta aventura. Gracias a Natalí, estos 2 años siempre tuvieron una sonrisa.



Índice general

1. Introducción	1
1.1. Problema de Investigación	1
1.2. Marco Conceptual	2
1.2.1. Elicitación de Requerimientos	2
1.2.2. Modelo de Casos de Uso	3
1.2.3. Historias de Usuario	4
1.2.4. Transparencia	4
1.3. Objetivo de Investigación	5
1.4. Antecedentes	5
1.5. Organización de la Tesis	7
2. Revisión Sistemática de la Literatura	9
2.1. Proceso de la Revisión	9
2.1.1. Planificación	9
2.1.2. Ejecución	11
2.2. Discusión	19
2.3. Amenazas a la Validez	20
2.3.1. Validez del Constructo	20
2.3.2. Validez Interna	20
2.3.3. Validez Externa	21
2.3.4. Validez de las Conclusiones	21
2.4. Conclusión de la Revisión Sistemática	21
3. Planteamiento del Experimento	23
3.1. Planificación del Experimento	23
3.1.1. Objetivos	23
3.1.2. Participantes	25
3.1.3. Instrumentos del Experimento	26
3.1.4. Hipótesis, Parámetros y Variables	27
3.2. Diseño del Experimento	29
3.2.1. Procedimiento	29
3.2.2. Validación del Experimento	30
3.2.3. Técnicas Estadísticas Aplicadas	30

Índice general

4. Evaluación	33
4.1. Ejecución del Experimento	33
4.1.1. Preparación	33
4.1.2. Desviaciones de la Planificación	34
4.2. Análisis de Resultados	34
4.2.1. Pregunta 1: Cantidad de consultas	35
4.2.2. Pregunta 2: Tiempo requerido para entender los documentos	38
4.2.3. Preguntas de Percepción (3, 4 y 5)	40
4.2.4. Análisis de las Relaciones del MEM	44
4.3. Discusión	47
4.3.1. Evaluación de Resultados e Implicancias	47
4.3.2. Amenazas a la Validez	49
4.4. Conclusiones de la Evaluación	50
5. Conclusiones	51
5.1. Revisión Sistemática	51
5.2. Evaluación Empírica	52
5.3. Trabajo Futuro	53
Anexos	61
A. Hoja Informativa	63
B. Formulario de Consentimiento	65
C. Cuestionario de Autoevaluación	69
D. Casos Propuestos	71
D.1. Caso 1(a)	71
D.2. Caso 1(b)	71
D.3. Caso 2(a)	72
D.4. Caso 2(b)	72
E. Registro de Historias de Usuario	73
F. Documento de Casos de Uso	75
G. Cuestionario de Percepción sobre el Análisis	79
H. Registro de Consultas	81
I. Diseño de Pantallas	83
J. Cuestionario de Percepción sobre el Diseño	85

Índice de tablas

2.1. Criterios de la técnica PICOC	10
2.2. PICOC para la Pregunta 1	10
2.3. PICOC para la Pregunta 2	11
2.4. PICOC para la Pregunta 3	11
2.5. Pasos de la búsqueda primaria	12
2.6. Fuentes electrónicas consultadas	12
2.7. Resultado de la búsqueda primaria	14
2.8. Número de artículos luego de aplicar los criterios de inclusión y exclusión	15
2.9. Publicaciones relevantes	15
2.10. Listado de preguntas para el aseguramiento de la calidad . .	16
2.11. Resultado de la evaluación de la calidad de los estudios selec- cionados	16
2.12. Conferencias y revistas relevantes	17
2.13. Características de los Modelos	18
3.1. PICOC para el diseño del experimento	24
3.2. Variables dependientes	28
4.1. Edad de los Participantes	35
4.2. Sexo de los Participantes	35
4.3. Selección de participantes de acuerdo a su experiencia en las etapas del ciclo de vida	35
4.4. Pregunta 1 - Estadísticos descriptivos para las consultas efec- tuadas durante el diseño	36
4.5. Pregunta 1 - Resultados de la prueba de bondad de ajuste . .	37
4.6. Pregunta 1 - Valores de los rangos resultantes de la prueba de U de Mann-Whitney	37
4.7. Pregunta 1 - Resultados de la prueba de U de Mann-Whitney	37
4.8. Pregunta 2 - Estadísticos descriptivos para las consultas efec- tuadas durante el diseño	38
4.9. Pregunta 2 - Resultados de la prueba de bondad de ajuste . .	38
4.10. Pregunta 2 - Resultados de la prueba de Levene	39

Índice de tablas

4.11. Pregunta 2 - Resultados de la prueba-t para 2 muestras independientes	39
4.12. Preguntas 3, 4 y 5 - Validez del cuestionario de percepción de análisis	41
4.13. Preguntas 3, 4 y 5 - Validez del cuestionario de percepción de diseño	41
4.14. Pregunta 3, 4 y 5 - Confiabilidad inicial de los cuestionarios de percepción	42
4.15. Preguntas 3, 4 y 5 - Estadísticos descriptivos para los cuestionarios de percepción de análisis y diseño	42
4.16. Preguntas 3, 4 y 5 - Resultados de la prueba de bondad de ajuste	43
4.17. Preguntas 3, 4 y 5 - Resultados de la prueba-t para 1 muestra con un $\mu = 3$ sobre los datos recabados a través del cuestionario de percepción del análisis	43
4.18. Preguntas 3 y 4 - Resultados de la prueba-t para 1 muestra con un $\mu = 3$ sobre los datos recabados a través del cuestionario de percepción del diseño	44
4.19. Pregunta 5 - Resultados de la prueba de Wilcoxon para 1 muestra con una mediana = 3 sobre los datos recabados a través del cuestionario de percepción del diseño	44
4.20. Correlaciones - Reglas propuestas por Muijis para determinar la fortaleza de la relación entre 2 variables	44
4.21. Correlaciones - Resumen de los datos recopilados de participantes con experiencia en análisis	45
4.22. Correlaciones - Descriptivos de los datos recopilados de participantes con experiencia en análisis	45
4.23. Correlaciones - Resultados de la prueba de bondad de ajuste de los datos recopilados de participantes con experiencia en análisis	45
4.24. Correlaciones - Resumen de correlaciones de los datos recopilados de participantes con experiencia en análisis	46
4.25. Correlaciones - Resumen de los datos recopilados de participantes con experiencia en diseño	46
4.26. Correlaciones - Descriptivos de los datos recopilados de participantes con experiencia en diseño	46
4.27. Correlaciones - Resultados de la prueba de bondad de ajuste de los datos recopilados de participantes con experiencia en diseño	47
4.28. Correlaciones - Resumen de correlaciones de los datos recopilados de participantes con experiencia en diseño	47

Índice de figuras

2.1. Protocolo de Revisión Sistemática Propuesto por Ahmad et al.	10
2.2. Publicaciones resultantes de la búsqueda primaria	13
2.3. Publicaciones al año 2015	17
2.4. Publicaciones según el tipo de documento	18
3.1. Modelo de Evaluación de Métodos propuesto por Moody et al.	25
4.1. Diagrama de cajas que representan la distribución de los datos asociados al total de las consultas efectuadas durante el diseño	36
4.2. Diagrama de cajas que representan la distribución de los datos asociados al tiempo que requirieron los participantes para comprender cada uno de los documentos de requerimientos .	39

Índice de figuras



CAPÍTULO 1

Introducción

El objetivo de este capítulo es brindar una introducción a la problemática que se busca afrontar a través de la presente tesis. Asimismo, se presenta un conjunto de conceptos que facilitarán la comprensión del tema expuesto en el documento.

El presente capítulo está conformado por 5 secciones. La Sección 1.1 presenta la problemática que esta investigación busca abordar, la Sección 1.2 presenta el marco conceptual, la Sección 1.3 presenta los objetivos del estudio, la Sección 1.4 presenta estudios que hayan abordado con anterioridad el uso de técnicas de elicitación de requerimientos para mejorar la comunicabilidad; finalmente la Sección 1.5 presenta la estructura del documento.

1.1. Problema de Investigación

La elicitación de requerimientos es una de las principales tareas a llevarse a cabo para la correcta implementación de un desarrollo de software. Su incorrecta especificación tiende a generar costos innecesarios a lo largo del proyecto e inclusive, su completo fracaso. De acuerdo a Boehm et al. [1], el 45% de los errores en el software tiene su origen en los requisitos y en el diseño preliminar; asimismo, la corrección de estos errores demanda un mayor esfuerzo que la corrección de errores generados durante la codificación. En 1995, el Standish Group [2] elaboró el Reporte Chaos en donde presentan que una de las 3 principales causas por las que un proyecto de software fracasa se debe a la poca claridad de los requerimientos; es decir, no están bien definidos o presentan ambigüedades.

Además, según Barbacci et al. [3], los requerimientos de software tienden a enumerar principalmente requerimientos funcionales dejando de lado o explicando muy ambiguamente los requerimientos no funcionales (RNF) asociados a los atributos de calidad del producto software. Esto trae como consecuencia que durante el diseño de la arquitectura del sistema, el arquitecto toma la responsabilidad de decidir sobre la prioridad de los atributos de calidad que deben de ser incluidos. Estas decisiones no necesariamente

concuerdan con las prioridades o necesidades de los interesados. De acuerdo a Nord et al. [4], en algunos proyectos, la tarea de diseño de la arquitectura no recibe la importancia debida, siendo omitida del ciclo de desarrollo del software.

1.2. Marco Conceptual

A continuación se describen un conjunto de conceptos que permitirán contextualizar las ideas presentadas en el presente estudio.

1.2.1. Elicitación de Requerimientos

De acuerdo a Bittner [5], los requerimientos de software son un conjunto de condiciones o capacidades que un sistema debe de cumplir. Asimismo, el SWEBOK [6] indica que cada uno de los requerimientos debe de ser verificable a nivel de funcionalidades o a nivel del sistema y se le debe de poder asignar una prioridad. Los requerimientos pueden definir tanto el producto a elaborar como el proceso que se seguirá para su elaboración.

Se cuentan con 2 categorías de requerimientos:

- **Requerimientos Funcionales:** Describen las funcionalidades que deben de estar implementadas en el software. También son conocidas como capacidades o funcionalidades que el producto software debe de cumplir.
- **Requerimientos No Funcionales:** Son las restricciones o requisitos de calidad del producto software. Están relacionados con requerimientos de desempeño, compatibilidad, usabilidad, fiabilidad, seguridad, mantenibilidad y portabilidad [7].

Todos los requerimientos elicitados deben de ser claros y no presentar ambigüedad [5] [6]; asimismo, deben de poder ser verificados dada la presencia de la funcionalidad - requisitos funcionales - o haciendo uso de valores cuantificables - requerimientos no funcionales.

La elicitación de requerimientos, también conocida como captura de requerimientos o adquisición de requerimientos, es el proceso que se preocupa por definir los orígenes de los requerimientos - a través de personas o de artefactos - y cómo el ingeniero de software los recabará. Es la primera etapa en el ciclo de desarrollo del software.

Uno de los principales factores para una adecuada elicitación, es una comunicación efectiva entre los distintos interesados del proyectos (patrocinador, usuarios, jefe de proyecto, equipo del proyecto, etc.). Las personas responsables del proceso de elicitar los requerimientos serán los responsables de mantener un canal de comunicación sin ruido entre los usuarios que

conocen el dominio del negocio y el equipo del proyecto quienes utilizan un lenguaje técnico, propio de la ingeniería de software.

1.2.2. Modelo de Casos de Uso

El modelo de casos de uso fue introducido por Ivar Jacobson en el año 1992. Representa un conjunto de escenarios organizados que se utilizan para definir el propósito de sistemas y software [8]. Su uso principal es el de ayudar en el proceso de captura de requerimientos y definir las interacciones entre el entorno (actores) y el sistema. Dentro de su descripción se incluyen las funcionalidades que el sistema debe de cumplir [9].

Cockburn [10] define los casos de uso como requerimientos que, si están bien definidos, pueden detallar el comportamiento deseado del software. Estos no requieren cubrir el íntegro del total de las funcionalidades listadas en el catálogo de requerimientos; sin embargo, deben de incluir las funcionalidades más relevantes. De acuerdo a Bittner [5], tradicionalmente cada caso de uso está conformado por los siguientes elementos:

- **Nombre del Caso de Uso:** Debe de ser único y resumir su resultado cuando un actor interactúa con él.
- **Breve descripción:** Una breve descripción del rol y propósito del caso de uso.
- **Flujo de eventos:** Describe el conjunto de acciones que intercambian el actor con el sistema. Esta descripción debe de ser de fácil entendimiento para los interesados del proyecto. Está conformado por un flujo primario, flujos alternativos y subflujos.
- **Requerimientos especiales:** Una descripción que recoge todos aquellos requerimientos que están asociados al caso de uso y que no figuran dentro del flujo de eventos; pero que deben de ser considerados para la correcta implementación del sistema. En esta sección se detallan, por ejemplo, los requerimientos no funcionales.
- **Precondiciones:** Descripción que presenta las restricciones que deben de cumplirse para que el caso de uso inicie.
- **Postcondiciones:** Descripción que representa las restricciones que deben de cumplirse cuando el caso de uso ha culminado.
- **Puntos de extensión:** Ubicaciones dentro del flujo de eventos del caso de uso en donde un comportamiento alternativo puede suscitarse.
- **Relaciones:** Las relaciones en las que el caso de uso participa.
- **Diagrama:** Diagramas que ilustran aspectos del caso de uso como la estructura del flujo de eventos o las relaciones que tiene el caso de uso. Por ejemplo, el diagrama de actividades.

1.2.3. Historias de Usuario

Las historias de usuarios fueron presentadas a través de la metodología Extreme Programming en el año 1998. Se caracterizan por ser historias cortas que describen alguna característica que el sistema a implementar debe cumplir [11], [12]. Las historias de usuario son comúnmente utilizadas en proyectos ágiles. Están conformadas por 3 elementos [11]:

- **La tarjeta:** Es una descripción de la historia en donde se presenta el rol del usuario, la tarea que desea llevar a cabo y el motivo que lo impulsa a llevar a cabo dicha acción.
- **La conversación:** Lugar en donde se detalla información adicional que complemente la descripción
- **La confirmación:** Es un conjunto de pruebas que sirve como indicador para saber cuando la historia ha sido completada.

Su principal beneficio es que se centran en la necesidades del usuario; y son cortas, concisas y fáciles de mantener a lo largo del proceso de desarrollo del producto software. De acuerdo a Hvalshagen et al. [13]; durante la última década los investigadores se han visto interesados en cómo las narrativas facilitan el proceso de aprendizaje y comprensión. Dentro de las razones por las que las historias son más fáciles se presentan:

- Proveen un contexto para el conocimiento nuevo.
- Permiten aplicar el conocimiento a problemas reales.

Esto es congruente con lo que indica Cockburn [10] sobre la tendencia de las personas a utilizar narrativas a fin de visualizar un escenario que emplea las funcionalidades a incluir en el software a implementar.

Entre las desventajas de su uso se encuentran la falta de detalle respecto al requerimiento a implementar; así como la falta de información respecto a las restricciones y los requisitos no funcionales del sistema como son los atributos de calidad.

1.2.4. Transparencia

Existen distintas definiciones del concepto de transparencia dependiendo del contexto en el cual el término es utilizado. Al efectuar una búsqueda en el diccionario en línea de la Real Academia Española, se encuentran las siguientes definiciones sobre el término “transparente”: *“Claro, evidente, que se comprende sin duda ni ambigüedad. Que se deja adivinar o vislumbrar sin declararse o manifestarse”*.

En esta tesis se utilizará la definición propuesta por Tu et al. [14] quien describe la transparencia como *el grado con el que los interesados en el proyecto pueden absolver sus dudas utilizando la información generada durante*

1.3. Objetivo de Investigación

el ciclo del vida del producto software. De acuerdo a la Guía del PMBOK [15], se entiende como interesado a cualquier persona o grupo de personas que podrían verse afectadas de forma positiva o negativa por una decisión, actividad o resultado de un proyecto.

La transparencia está conformada por 3 atributos fundamentales:

1. **Accesibilidad** a la información por parte del interesado de forma sencilla;
2. **Entendibilidad** de la información accedida por los interesados; y
3. **Relevancia** de la información accedida por los interesados para absolver sus preguntas.

1.3. Objetivo de Investigación

El objetivo de este estudio es el de evaluar el impacto en la transparencia de los requerimientos levantados con el apoyo de 2 técnicas de elicitación conocidas: los modelos de casos de uso y las historias de usuario. Para poder cubrir el objetivo del estudio, se definieron los siguientes hitos.

- Efectuar una revisión del estado actual de los modelos propuestos que hagan un uso combinado de las historias de usuario y casos de uso.
- Identificar cuáles de los modelos propuestos cuentan con herramientas que les permite trabajar adecuadamente con requerimientos no funcionales y cuáles han sido evaluados dentro de la industria.
- Efectuar un experimento empírico sobre una propuesta de solución que combina las herramientas: los modelos de casos de uso y las historias de usuario.

1.4. Antecedentes

A la fecha existen otros estudios que han abordado la problemática asociada a la poca comprensibilidad de los requerimientos recabados durante elicitación. Asimismo, algunos estudios presentan soluciones propuestas que integran el uso de las técnicas de historias de usuario y los casos de uso. Sin embargo, en ninguno de los trabajos encontrados, se estudia la posibilidad de utilizar ambas técnicas para mejorar la transparencia en la comunicabilidad y la comprensibilidad de los requerimientos.

Un experimento del 2009 elaborado por España et al. [16], propone la inclusión de técnicas de Análisis de la Comunicación en el modelo de Casos de Uso ya que permitiría elicitar requerimientos con un mayor grado de completitud a nivel del encapsulamiento funcional y en la comunicabilidad.

Capítulo 1. Introducción

Por otro lado, el estudio de Abrahão et al. [17] del 2011 propone una extensión a la metodología RUP para mejorar la efectividad en la elicitación de requerimientos. Para medir el impacto real de la técnica y la percepción de los participantes en su uso, utilizaron el *Method Evaluation Model* - MEM - propuesta por Moody et al. [18]. El estudio se basó en las suposiciones que las técnicas de elicitación de requerimientos (i) deben de facilitar la comunicación de las necesidades entre el interesado y los desarrolladores; y (ii) deben de brindar detalles específicos de las tareas que el software debe de poder llevar a cabo.

En otros estudios se llevan a cabo comparaciones a nivel de comunicabilidad y transparencia entre 2 técnicas. En el 2012, Wanderley y Silva [19] reportaron que en la comunicación entre los interesados en el proyecto y el equipo de desarrollo tiende a haber pérdida de la información. Esta pérdida termina generando una brecha en la semántica entre lo que se desea implementar y lo que es realmente producido. Esta brecha es evidenciada por las inconsistencias presentes en los artefactos generados durante la elicitación de requerimientos y el modelo conceptual de la información. En el 2014 [20], se efectuó un experimento en donde se comparaba la reducción de la pérdida de información al utilizar un modelo basado en mapas conceptuales contra los diagramas de clase de UML. Los resultados preliminares de este estudio presentan que la técnica basada en mapas conceptuales permiten entender y comprender el dominio del negocio con mayor facilidad.

Asimismo, Hadar et al. [21] en el 2013 efectuó la comparación bajo las técnicas de Casos de Uso y Tropos desde el punto de vista del análisis de requerimientos. Los autores determinaron que al utilizar Tropos, los requerimientos elicitados eran más comprensibles que cuando se utilizaban Casos de Uso; sin embargo, el uso de Tropos requería un mayor tiempo que al aplicar su contraparte.

En el 2015, Tu et al. [22] presentó un estudio en donde comparaba a nivel de eficiencia, efectividad y satisfacción la comprensibilidad de los requerimientos funcionales presentados en un documento de especificación de requerimientos contra un modelado de casos de uso. La autora concluye que el uso de los casos de uso presentados en forma narrativa brinda una mayor transparencia en los requerimientos del software y ello mejora la comunicabilidad entre los analistas y los interesados. Sin embargo, el estudio elaborado utiliza como interesados a personas que se desempeñan dentro del campo de la ingeniería informática lo cual genera un sesgo en el estudio debido a que dentro de las organizaciones, los interesados por lo general no cuentan con conocimientos en informática o en ingeniería de software.

Finalmente, se han encontrado algunos estudios como los de Gallardo-Valencia et al. [9] y Hvalshagen et al. [13] que han evaluado la combinación de historias de usuarios y casos de uso; mas no han profundizado en temas asociados al grado de comprensibilidad de los mismos. No obstante, presentan un significativo aporte en este campo de estudio que puede ser utilizado

para presentar una nueva propuesta de solución o para nutrir a alguno de los estudios existentes.

1.5. Organización de la Tesis

La organización del trabajo de tesis es la siguiente:

- Capítulo 2: Presenta una revisión sistemática de la literatura siguiendo el protocolo propuesto por Ahmad et al. [23]. Esta revisión refleja el estado del arte en torno a modelos propuestos que hacen un uso combinado de historias de usuario y casos de uso.
- Capítulo 3: Presenta el objetivo del experimento empírico, las preguntas de investigación que se buscan responder y las hipótesis asociadas a cada una. Asimismo, brinda una descripción del diseño del experimento.
- Capítulo 4: Presenta las tareas que se llevaron a cabo durante la ejecución del experimento, la descripción de los datos recabados y los resultados obtenidos del análisis estadístico de los datos.
- Capítulo 5: Presenta las conclusiones de la tesis; así como futuros estudios propuestos.
- Anexos: Los anexos contienen el diseño de los artefactos utilizados durante la ejecución del experimento.
 - Anexo A. Hoja Informativa
 - Anexo B. Formulario de Consentimiento
 - Anexo C. Cuestionario de Autoevaluación
 - Anexo D. Casos Propuestos
 - Anexo E. Registro de Historias de Usuario
 - Anexo F. Documento de Casos de Uso
 - Anexo G. Cuestionario de Percepción sobre el Análisis
 - Anexo H. Registro de Consultas
 - Anexo I. Formulario de Diseño
 - Anexo J. Cuestionario de Percepción sobre el Diseño

Capítulo 1. Introducción



CAPÍTULO 2

Revisión Sistemática de la Literatura

Las Revisiones Sistemáticas de la Literatura (*Systematic Literature Review - SLR*) brindan información del estado del arte de la temática a investigar. Previamente a efectuar un trabajo de investigación, es recomendable verificar la existencia de SLR a fin de contar con un panorama general del estado del arte del tema a investigar; siendo necesaria la elaboración de una SLR en caso no se encontrara alguna que cubra los objetivos de la investigación.

La organización del presente capítulo se detalla a continuación. La Sección 2.1 describe el protocolo seguido para la revisión. La Sección 2.2 presenta la discusión en base a la información recabada al aplicar el protocolo. La Sección 2.3 presenta las amenazas a la validez de la revisión. Finalmente, en la Sección 2.4 se resume las conclusiones derivadas de la discusión.

2.1. Proceso de la Revisión

Para el presente estudio se ha optado por elaborar una revisión sistemática de literatura (Systematic Literature Review, SLR) que, según Kitchenham et al. [24], es un mecanismo que permite identificar, evaluar e interpretar un conjunto de investigaciones enfocadas en un tema en particular. El objetivo de este trabajo es la identificación y revisión de los modelos de elicitación de requerimientos de software que combinen técnicas de casos de uso e historias de usuario. Para el proceso de revisión se ha seguido el protocolo propuesto por Ahmad et al. [23] representado en la Figura 2.1.

2.1.1. Planificación

Preguntas de Investigación

A fin de poder entender los resultados obtenidos, se hizo uso de cinco criterios: Población, Intervención, Comparación, Resultados y Contexto (PICOC); los mismos que se encuentran detallados en la Tabla 2.1.

Capítulo 2. Revisión Sistemática de la Literatura

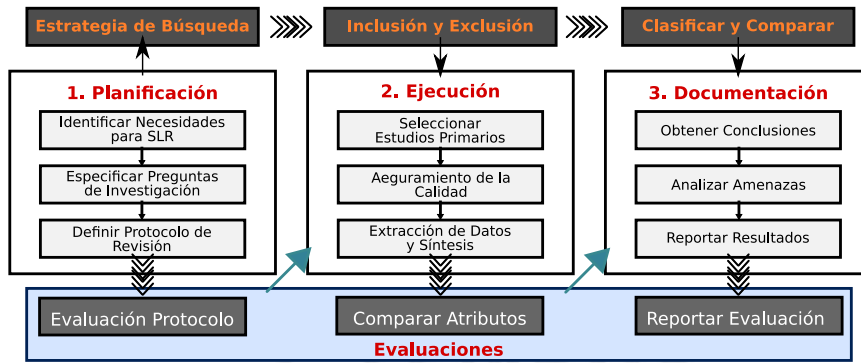


Figura 2.1: Protocolo de Revisión Sistemática Propuesto por Ahmad et al. [23]

Tabla 2.1: Criterios de la técnica PICOC

Criterio	Significado
Población	¿Quién o Qué?
Intervención	¿Cómo?
Comparación	¿Cuál es la alternativa con la que se está comparando?
Resultados	¿Qué estamos tratando de obtener, mejorar, proponer?
Contexto	¿Bajo qué circunstancias?

En base a la temática inicialmente planteada, se han elaborado las siguientes preguntas de investigación.

Pregunta 1: ¿Actualmente existen modelos o propuestas de modelos que combinen casos de uso e historias de usuario?

La construcción de esta pregunta se ha basado en los elementos de Población e Intervención (PI) de la técnica PICOC. Tal como se aprecia en la Tabla 2.2, el objetivo de esta pregunta es obtener un estado del arte del total de publicaciones en la población de ingeniería de software o requisitos de software que hagan mención de casos de uso e historias de usuario.

Tabla 2.2: PICOC para la Pregunta 1

Criterio	Alcance
Población	Ingeniería de Software o Requisitos de Software o Desarrollo de Software
Intervención	Técnicas que empleen Casos de Uso e Historias de Usuario

Pregunta 2: De los modelos existentes o propuestos, ¿alguno de ellos toman en consideración los requerimientos no funcionales?

En base a los resultados previamente obtenidos, esta pregunta busca determinar cuáles son los modelos que toman en consideración el levantamiento de los requerimientos no funcionales del proyecto. En la Tabla 2.3 se muestra la estructura de esta pregunta.

2.1. Proceso de la Revisión

Tabla 2.3: PICOC para la Pregunta 2

Criterio	Alcance
Población	Ingeniería de Software o Requisitos de Software o Desarrollo de Software
Intervención	Técnicas que empleen Casos de Uso e Historias de Usuario
Resultados	Levantamiento de requisitos no funcionales

Pregunta 3: De los modelos existentes o propuestos, ¿alguno de ellos ha sido validado o es actualmente utilizado en la industria?

Esta pregunta busca acotar los resultados obtenidos de la Pregunta 1 basándose en la clasificación propuesta por Wieringa et al. [25] en donde las publicaciones se agrupan en 6 categorías (evaluación, propuesta de solución, validación, filosofía, opinión y experiencias personales). La respuesta a esta pregunta permitirá conocer si el modelo propuesto ha sido validado dentro de un entorno académico o evaluado tras su aplicación en la industria. La Tabla 2.4 muestra como ha sido estructurada esta pregunta.

Tabla 2.4: PICOC para la Pregunta 3

Criterio	Alcance
Población	Ingeniería de Software o Requisitos de Software o Desarrollo de Software
Intervención	Técnicas que empleen Casos de Uso e Historias de Usuario
Contexto	Entorno académico o aplicación en la industria

2.1.2. Ejecución

Selección de los Estudios

Sobre el conjunto de publicaciones a obtenerse durante las búsquedas primarias y secundarias, se aplicarán los siguientes 2 criterios de selección.

Criterios de Inclusión

- Las publicaciones deben de estar en inglés.
- Se debe de contar con acceso al contenido de la publicación.
- Los trabajos publicados deben de ser revisiones, propuestas de solución, validaciones o evaluaciones.

Criterios de Exclusión

- Las publicaciones que no propongan un modelo que combine ambas técnicas - casos de uso e historias de usuario.

Capítulo 2. Revisión Sistemática de la Literatura

- Las publicaciones que no validen o evalúen un uso combinado de ambas técnicas - casos de uso e historias de usuario.
- Los artículos de opinión.
- Las publicaciones duplicadas. En caso hubieran resultados duplicados, se considerará aquellas cuya información sea más completa.

Estrategias de Búsqueda

El proceso de búsqueda estuvo conformado por una búsqueda primaria y una secundaria.

Búsqueda Primaria

Se hicieron uso de las 5 etapas propuestas por Ahmad et al. [23]; éstas han sido detalladas en la Tabla 2.5.

Tabla 2.5: Pasos de la búsqueda primaria

Pasos	Descripción
Elaborar las cadenas de búsqueda	Se elaboraron en base a los términos presentados en las Tablas 2.2, 2.3 y 2.4
Considerar sinónimos	Considerar sinónimos para los términos utilizados en la búsqueda.
Combinar los términos de búsqueda	Hacer uso del conector OR para combinar sinónimos y el conector AND para interconectar los principales términos de búsqueda.
Dividir las cadenas de búsqueda	Dividir la cadena en varias subcadenas a fin de que puedan ser aplicadas a las distintas fuentes de datos.
Gestionar las referencias	Se hizo uso de la herramienta Mendeley

Las siguientes cadenas fueron ejecutadas en las fuentes de datos especificadas en la Tabla 2.6 durante el mes de abril de 2015. Las búsquedas se realizaron sobre los títulos, resúmenes y palabras claves de las publicaciones de acuerdo a las opciones que brinda cada una de las fuentes digitales. Asimismo, no se aplicó ninguna restricción basada en la fecha de publicación del trabajo.

Tabla 2.6: Fuentes electrónicas consultadas

Id	Base de Datos	URL
SS	Sciverse Scopus	http://scopus.com/
IEEE	IEEEExplore	http://ieeexplore.ieee.org
ACM	ACM Digital Library	http://dl.acm.org/
ISI	ISI Web of Science	http://isiknowledge.com

2.1. Proceso de la Revisión

- **Cadena 1:** ((“user story” OR “user stories”) AND (“use case” OR “use cases”)) AND (“software development” OR “software construction” OR “software project” OR “software projects” OR “software process” OR “software processes” OR “software engineering” OR “requirement* engineering” OR “requirement*”)¹
- **Cadena 2:** Cadena 1 AND (“Non Functional Requirement*” OR NFR)²
- **Cadena 3:** Cadena 1 AND (study OR studies OR experiment* OR Verificat* OR Validat* OR Evaluation*)³

El resultado de la búsqueda primaria retornó un total de 42 artículos publicados entre los años 2003 y 2014, de los cuales 16 eran duplicados. El listado de estudios únicos se encuentra enumerado en la Tabla 2.7.

Asimismo, como se puede apreciar en la Figura 2.2, la fuente electrónica que retornó un mayor número de resultados fue Sciverse Scopus con 19 publicaciones; seguido por ACM Digital Library e ISI Web of Science, cada una con 8 estudios.

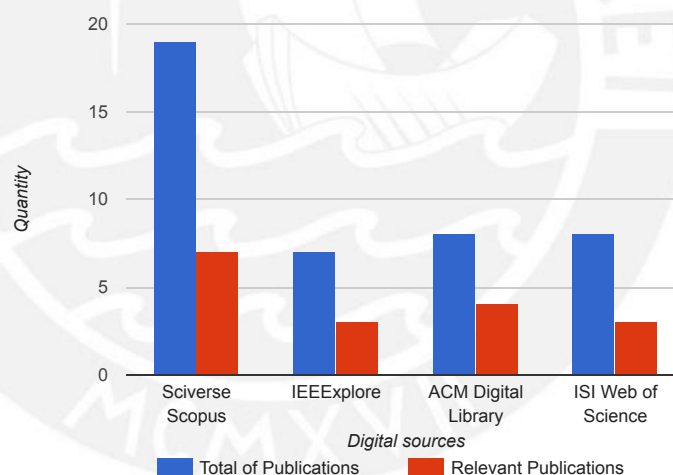


Figura 2.2: Publicaciones resultantes de la búsqueda primaria

Sobre las 26 publicaciones únicas, se aplicaron los criterios de selección a sus resúmenes. En caso que el resumen del artículo no incluyera suficiente información para tomar una decisión, se revisaron la introducción y conclusiones del mismo. Luego de ello quedaron 15 publicaciones de las que fueron seleccionadas 8 tras haber realizado una revisión de texto completo.

Búsqueda Secundaria

En la búsqueda secundaria se revisaron:

¹Esta cadena ha sido derivada de la Tabla 2.2.

²Esta cadena ha sido derivada de la Tabla 2.3.

³Esta cadena ha sido derivada de la Tabla 2.4.

Tabla 2.7: Resultado de la búsqueda primaria

Ref.	Fuente Electrónica	Cadenas
[26]	SS	1
[27]	SS	1
[28]	SS	1
[29]	ACM	1
[30]	SS, ISI	1, 2
[9]	SS, IEEE, ACM	1, 3
[31]	SS, ACM, ISI	1, 3
[32]	SS	1
[33]	SS	1
[34]	SS, ISI	1
[35]	IEEE	1
[36]	SS, IEEE, ISI	1, 2
[37]	SS, IEEE, ISI	1, 2
[38]	SS	1
[39]	IEEE	1
[40]	SS	1, 3
[41]	IEEE	1
[42]	SS	1
[43]	SS, ACM, ISI	1
[44]	IEEE	1
[45]	ACM	1, 2
[46]	SS, ACM, ISI	1
[47]	SS, ACM	1
[13]	SS	1, 3
[11]	ACM	1
[48]	SS, ISI	1

- Los trabajos que citen a las publicaciones obtenidas en la búsqueda primaria.
- Los trabajos que sean citados por las publicaciones resultantes de la búsqueda primaria.

Asimismo, las publicaciones seleccionadas deben de cumplir con los criterios de inclusión y exclusión previamente definidos.

El total de publicaciones citadas por los 8 artículos seleccionados de la búsqueda primaria, o que citaban a alguno de ellos, fue 221 (191 sin duplicados). Luego de filtrar los resúmenes, el total de artículos fue reducido a 11. Finalmente tras revisar el texto completo de las publicaciones restantes, se obtuvieron las siguientes 3 publicaciones: [49], [28] y [50].

El proceso de selección de los artículos relevantes es resumido en la Tabla 2.8.

Finalmente, el detalle de las publicaciones relevantes obtenidas del proceso de búsqueda se detallan en la Tabla 2.9.

2.1. Proceso de la Revisión

Tabla 2.8: Número de artículos luego de aplicar los criterios de inclusión y exclusión

	Resultados	Sin Duplicados	Criterios en el Resumen	Revisión Texto Completo
Primaria	42	26	15	8
Secundaria	221	191	11	3

Tabla 2.9: Publicaciones relevantes

Ref.	Fuente	Conferencias y Revistas	Modelo	Tipo [25]	RNF
[9]	SS, IEEE, ACM	CERE	-	Validación	No
[31]	SS, ACM, ISI	Computers in Industry	Athena	Propuesta Validación	Sí
[50]	-	REFSQ	-	Validación Evaluación	No
[36]	SS, IEEE, ISI	IEEE SOUTHEAST-CON	NORMAP	Propuesta Validación	Sí
[37]	SS, IEEE, ISI	IEEE SOUTHEAST-CON	NORMAP	Propuesta Validación	Sí
[40]	SS	WEASELTech	-	Validación	No
[49]	-	APSEC	NORMAP	Propuesta Validación	Sí
[45]	ACM	Nova Southeastern University	NORMAP	Propuesta	Sí
[47]	SS, ACM	ICSE	K-gileRE	Propuesta Validación	No
[13]	SS	AMCIS	-	Filosófico Validación	No
[51]	-	IEEE SOUTHEAST-CON	NORMAP	Propuesta Validación	Sí

Evaluación de la Calidad del Estudio

Los 11 estudios previamente seleccionados fueron sometidos a un proceso de evaluación de la calidad. En la Tabla 2.11 se puede observar los resultados de la evaluación de calidad luego de haber aplicado el instrumento de evaluación presentado en la Tabla 2.10, que está basado en la propuesta de Zarour et al. [52]. Este instrumento utiliza una escala de calificación de 3 niveles (Sí = 1 punto, No = 0 punto, Parcialmente = 0.5 punto).

Todas aquellas publicaciones relevantes con una calificación igual o me-

Tabla 2.10: Listado de preguntas para el aseguramiento de la calidad

ID	Pregunta
QA1	¿El objetivo de la investigación ha sido suficientemente explicado?
QA2	¿La idea o el enfoque presentado ha sido claramente explicado?
QA3	¿Se han considerado amenazas a la validez?
QA4	¿Hay una adecuada descripción del contexto en donde se ha llevado a cabo la investigación?
QA5	¿Se mencionan de forma clara los hallazgos del estudio?

nor a 5.0 y mayor o igual a 2.5 son consideradas como aceptadas; mientras que aquellas con un puntaje menor a 2.5 son descartadas. Los puntajes alcanzados por cada estudio, no son menores a 3.0; asimismo, el puntaje promedio es de 3.95. El resultado de la evaluación de la calidad demuestra que los 11 estudios son aceptables.

Tabla 2.11: Resultado de la evaluación de la calidad de los estudios seleccionados

Ref.	QA1	QA2	QA3	QA4	QA5	Calificación
[9]	1.0	1.0	1.0	1.0	1.0	5.0
[31]	1.0	1.0	0.5	1.0	0.5	4.0
[50]	1.0	1.0	1.0	1.0	1.0	5.0
[36]	1.0	1.0	0.0	1.0	1.0	4.0
[37]	1.0	1.0	0.0	1.0	1.0	4.0
[40]	1.0	1.0	0.0	1.0	0.5	3.5
[49]	1.0	0.5	0.0	1.0	1.0	3.5
[45]	1.0	1.0	0.0	1.0	1.0	4.0
[47]	1.0	1.0	0.5	1.0	0.5	4.0
[13]	1.0	1.0	0.0	0.0	1.0	3.0
[51]	1.0	1.0	0.0	1.0	0.5	3.5

Extracción de Datos y Síntesis

Las búsquedas se efectuaron durante el mes de abril del 2015 sobre fuentes digitales presentadas en la Tabla 2.6 y luego de aplicar los criterios de selección, se obtuvieron el listado de publicaciones de acuerdo a la Tabla 2.9.

A continuación se procederá a extraer y sintetizar los datos provistos por cada una de las publicaciones.

Distribución de las publicaciones por año

La Fig. 2.3 es una gráfica de barras que muestra en su eje vertical, el total de publicaciones por cada año representado en el eje horizontal. El rango de fechas de las publicaciones abarca desde el año 2007 hasta el 2015. Dentro de este período de tiempo, el año 2012 es el que tuvo un mayor número de publicaciones con 3 trabajos, seguido por el 2007 con 2 trabajos.

2.1. Proceso de la Revisión

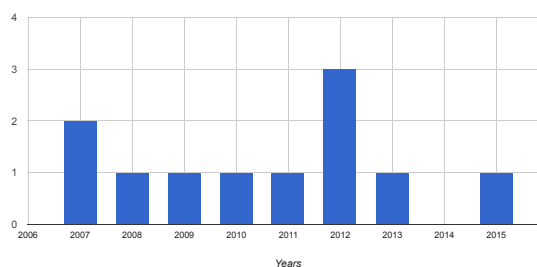


Figura 2.3: Publicaciones al año 2015

Conferencias y revistas relevantes

La Tabla 2.12 muestra el listado de conferencias y revistas encontradas en base al análisis cualitativo de los resultados. De acuerdo a los estudios seleccionados, IEEE SOUTHEASTCON es la fuente en donde se han presentado un mayor número de publicaciones asociadas al tema de la presente revisión.

Tabla 2.12: Conferencias y revistas relevantes

Nombre de la Fuente	Tipo	Cantidad
IEEE SOUTHEASTCON	Conferencia	3
AMCIS - Americas Conference on Information Systems	Conferencia	1
APSEC - Asia-Pacific Software Engineering Conference	Conferencia	1
CERE - International Workshops on Comparative Evaluation in Requirements Engineering	Conferencia	1
ICSE - International Conference on Software Engineering	Conferencia	1
LNCS - Lecture Notes in Computer Science	Conferencia	1
WEASELTech - International Workshop on Empirical Assessment of Software Engineering Languages and Technologies	Conferencia	1
Computers in Industry	Revista	1
Nova Southeastern University	Tesis	1

Además, como puede visualizarse en la Figura 2.4, el 81.8% de las publicaciones han sido presentadas en Conferencias, el resto de trabajos, se divide de forma equitativa entre revistas y tesis.

Síntesis de las publicaciones

De los trabajos seleccionados, se identificaron 3 metodologías para la elicitación de requerimientos que efectúan un uso combinado de historias de usuario y casos de uso. La Tabla 2.13 detalla el listado de características para cada una de las metodologías.

Capítulo 2. Revisión Sistemática de la Literatura

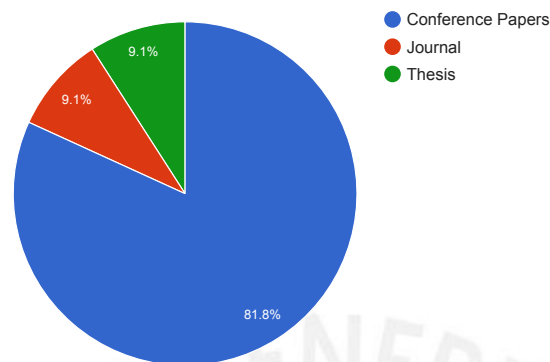


Figura 2.4: Publicaciones según el tipo de documento

Tabla 2.13: Características de los Modelos

Característica	Athena	K-gileRE	NORMAP
Aplica las US previamente a los UC	X	X	X
Considera la negociación de los requerimientos	X	-	-
Cuenta con una herramienta CASE	X	X	X
Efectúa un análisis de riesgos	-	-	X
El conocimiento es compartido entre los analistas e interesados	X	X	-
El modelo ha sido Evaluado	-	-	-
El modelo ha sido Validado	X	X	X
Incorpora Trazabilidad	X	-	-
Requiere contar con métricas de gestión y calidad preestablecidas	-	-	X
Requiere de un entorno para gestionar el conocimiento	X	X	-
Requiere que los interesados cumplan con roles distintos	X	X	-
Requiere una comunicación colectiva con interesados	X	X	-
Soporta Requerimientos No Funcionales	X	-	X
Uso de operadores punteros Orientados a Aspecto	-	-	X
Utiliza “Activity Theory” para la especificación de UC	X	-	-
Utiliza adicionalmente Escenarios	X	-	-
Utiliza algún esquema de tablas de conversión	X	-	X
Utiliza la técnica de “Group Story Telling”	X	-	-
Utiliza ontologías	-	X	-

En base a la información resumida en la Tabla 2.13, se ha encontrado la siguiente evidencia. Las propuestas identificadas hacen un uso combinado de casos de uso e historias de usuario. Las historias de usuario son aplicadas previamente a la elaboración de los casos de uso. Sin embargo, el proceso de la elaboración de las historias de usuario es distinto en cada una de las técnicas. Mientras que en Athena y K-gileRE las historias son elaboradas por la comunicación colectiva entre los interesados y los analistas, en NORMAP cada uno de los requerimientos del catálogo es convertido a una historia de usuario haciendo uso de las tarjetas con formato W^8 , las cuales extienden la información de las historias de usuario.

Otro aspecto importante a detallar es que a pesar de que en el listado de características del modelo Athena figura su capacidad para la elicitación de los requerimientos no funcionales, dicha funcionalidad se ve limitada al nivel de detalle con el que han sido declaradas las historias de usuario; si no incluyen suficiente información que permita la identificación de un requerimiento no funcional, el modelo no lo identificará. En contraste, NORMAP cuenta con métricas preestablecidas que permiten la adecuada detección de requerimientos no funcionales y el impacto que presentan dentro del análisis del riesgo de los proyectos.

2.2. Discusión

A fin de responder a las preguntas de investigación que han sido definidas en la Sección III, esta sección describe los hallazgos basados en las fuentes seleccionadas según la Tabla 2.9.

Pregunta 1: ¿Actualmente existen modelos o propuestas de modelos que combinen casos de uso e historias de usuario?

Respuesta 1: En base a los resultados obtenidos de la revisión sistemática, es posible determinar que al menos existen 3 modelos de elicitación de requerimientos de software que combinan casos de uso e historias de usuario. De los hallazgos se puede listar los modelos Athena propuesto por Laporti et al. [31], K-gileRE propuesto por Kumar et al. [47] y NORMAP propuesto por Farid et al. [49], [36], [37], [51], [45]

Pregunta 2: De los modelos existentes o propuestos, ¿alguno de ellos toma en consideración los requerimientos no funcionales?

Respuesta 2: De los modelos presentados en la Respuesta 1, NORMAP [45] es un modelo que hace uso de una extensión de las historias de usuario que busca recabar información adicional sobre las necesidades del cliente, como priorización de los atributos de calidad del software a desarrollar y un análisis del impacto que tiene sobre otros requerimientos [36], [51]. Este modelo presenta 3 componentes: AUC (Casos de Uso simplificados), ALC (Requerimiento No Funcional asociado al AUC) y ACC (Propuesta de solución asociada a un ALC) [49], [37].

Capítulo 2. Revisión Sistemática de la Literatura

En el caso del modelo Athena, los autores mencionan que el modelo está preparado para trabajar con requerimientos no funcionales; sin embargo, este modelo no presenta las herramientas necesarias para garantizar la elicitación de este tipo de requerimientos [31].

Pregunta 3: De los modelos existentes o propuestos, ¿alguno de ellos ha sido validado o es actualmente utilizado en la industria?

Respuesta 3: De los modelos resultantes de la búsqueda en la Respuesta 1, el modelo Athena ha sido validado en entornos controlados; asimismo los autores planean evaluar el desempeño de este modelo en un escenario real en una siguiente etapa [31]. En el caso de K-gileRE, el marco de trabajo ha sido validado dentro de un proyecto pequeño en el que ha presentado buenos resultados; sin embargo, como indican los autores, es necesario efectuar una nueva validación con proyectos de mayor tamaño a fin de poder predecir el comportamiento del marco en proyectos reales [47].

Adicionalmente, a pesar de que los artículos de Gallardo Valencia et al. [9] [40] y de Hvalshagen et al. [13] no proponen un modelo definitivo; la serie de experimentos que han realizado a fin de determinar la viabilidad del uso combinado de casos de uso e historias de usuario presentan un significativo aporte en este campo de estudio que puede ser utilizado para presentar una nueva propuesta de solución o para nutrir a alguno de los modelos existentes.

2.3. Amenazas a la Validez

De acuerdo a Jedlitschka et al. [53] todas aquellas amenazas que puedan impactar en los resultados obtenidos han de ser discutidas. Se identifican 4 posibles amenazas a la validez.

2.3.1. Validez del Constructo

Los principales constructos utilizados en este estudio son las cadenas de búsqueda elaboradas para cada una de las preguntas de investigación. Estas cadenas de búsqueda fueron construidas haciendo uso de sinónimos asociados a los términos que formaban parte del PICOC de cada una de las preguntas. Asimismo, cada una de estas cadenas ha sido ejecutada sobre un conjunto de fuentes digitales previamente seleccionadas. Existe la posibilidad de que algunos estudios relevantes no hayan sido considerados en caso estuvieran indexados en una de las fuentes digitales no revisadas o en caso contuvieran términos ajenos a los seleccionados para construir las cadenas de búsqueda.

2.3.2. Validez Interna

Es posible que se introdujera un sesgo al extraer y analizar los datos del estudio elaborado por el investigador. A fin de mitigar este riesgo, se

2.4. Conclusión de la Revisión Sistemática

incluyeron actividades para validar el análisis que fueron ejecutadas por el asesor del estudio.

2.3.3. Validez Externa

Está asociada con la capacidad para poder generalizar los resultados obtenidos de este estudio. A fin de mitigar este riesgo, el proceso de búsqueda es efectuado varias veces.

2.3.4. Validez de las Conclusiones

Es posible que algunos de los estudios fueran incorrectamente excluidos en esta revisión. A fin de mitigar este riesgo, se llevó una construcción cuidadosa de los criterios de inclusión y exclusión y una posterior validación de la calidad de los estudios seleccionados a fin de evitar la exclusión de estudios relevantes.

2.4. Conclusión de la Revisión Sistemática

El objetivo de esta revisión es el efectuar una evaluación de modelos que combinen 2 de las herramientas más utilizadas en este ámbito: los modelos de casos de uso y las historias de usuario. Para ello, se elaboraron 3 preguntas de investigación que fueron respondidas en base a los datos recabados en el proceso de búsqueda.

En la revisión sistemática efectuada, se identificaron un total de 11 artículos relevantes - conformado por los resultados de las búsquedas primaria y secundaria. De estos trabajos, se identificaron 3 propuestas metodológicas: Athena, K-gileRE y NORMAP.

Finalmente, se concluye que aunque las 3 propuestas han sido validadas en entornos controlados, ninguna de ellas ha sido evaluada en un entorno no controlado por lo que no es posible determinar su eficacia en el ámbito industrial.

Además, los 3 métodos presentan algunas características y asunciones que no necesariamente corresponden con su aplicación en un entorno de trabajo real. En el caso de Athena y K-gileRE, los autores asumen que los interesados contarán con la disponibilidad y el interés de trabajar de forma colectiva entre ellos y con los analistas; así como la predisposición de los analistas a compartir el conocimiento - del análisis de los requisitos - con los interesados. No existen garantías que aseguren el cumplimiento de esas asunciones debido a que no se puede afirmar que las personas participantes en el proyecto compartan una misma ideología sobre como llevar a cabo su trabajo.

En el caso del método NORMAP, el modelo ha sido pensado para su aplicación en proyectos que hacen uso de metodologías ágiles; mas no hay

Capítulo 2. Revisión Sistemática de la Literatura

información y validaciones efectuadas sobre su eficiencia en proyectos que no trabajan con dicha metodología. Cabe mencionar que a pesar de que NORMAP ha sido construido para ser aplicado en proyectos ágiles, la alta complejidad del mismo genera una fuerte dependencia con su herramienta CASE (NORMATIC).

Asimismo, los AUC que este modelo de trabajo propone brindan información visual asociada a los RNF; no obstante, se excluye la información de los actores y su relación con los casos de uso. Como consecuencia, con tan solo ver el diagrama generado no es posible determinar el actor que inicializa cada uno de los casos de uso.

Dado el poco número de estudios que evalúan la efectividad del uso combinado de historias de usuario y casos de uso, la siguiente etapa de este trabajo es evaluar la transparencia durante la elicitación de requerimientos producto del uso integrado de las historias de usuario y los casos de uso.



CAPÍTULO 3

Planteamiento del Experimento

El objetivo de este capítulo es presentar el conjunto de decisiones consideradas al elaborar la planificación de un experimento que permite evaluar el impacto en la transparencia de los requerimientos elicitados al combinar las técnicas de historias de usuarios y el modelo de casos de uso.

El capítulo se encuentra organizado de acuerdo a las guías propuestas por Jedlitschka et al. [53] y que a la fecha ha sido citada por más de 130 estudios. La Sección 3.1 lista las tareas efectuadas durante la planificación del experimento. La Sección 3.2 comprende el diseño del experimento.

3.1. Planificación del Experimento

A fin de poder establecer los elementos sobre los que se llevará a cabo el experimento, se utilizó la técnica PICOC⁴, que establece los siguientes criterios: Población, Intervención, Comparación, Resultados y Contexto. En base a esta técnica, se procedió con el diseño del experimento. Las características se resumen en la Tabla 3.1

3.1.1. Objetivos

El principal objetivo de este experimento empírico es evaluar una propuesta de solución que integre las técnicas de modelos de casos de uso e historias de usuario para determinar la transparencia en la comunicación efectiva entre los interesados en el proyecto y el equipo de desarrollo.

De acuerdo a Abrahão et al. [17] es posible adaptar el método MEM - *Method Evaluation Model* - propuesto por Moody et al. [18] para ser utilizado en el estudio de técnicas de elicitación de requerimientos. Este método de evaluación no solamente analiza la efectividad real de la técnica estudiada; sino también la efectividad percibida. La Figura 3.1 muestra el constructo del modelo utilizado para el presente estudio que abarca las variables asociadas a la eficacia real y a la eficacia percibida.

⁴Ver Tabla 2.1

Capítulo 3. Planteamiento del Experimento

Tabla 3.1: PICOC para el diseño del experimento

Criterio	Elemento
Población	Alumnos de los últimos ciclos de la carrera de ingeniería informática que hayan realizado prácticas profesionales.
Intervención	Evaluarán el marco de trabajo propuesto que integra casos de uso e historias de usuario.
Comparación	Se comparará contra el uso independiente de las historias de usuario.
Resultados	Se espera obtener que el uso combinado de ambas técnicas sea más efectivo, eficiente y satisfactorio.
Contexto	El experimento será llevado a cabo bajo la premisa que el interesado en el proyecto no es una persona afín a la carrera de ingeniería informática.

En MEM, se define la eficacia como el grado de efectividad y eficiencia con el que un método alcanza sus objetivos. Asimismo, los constructos del MEM se basan en el modelo TAM - *Technology Acceptance Model* propuesto por Davis [54] que ha sido validado para la evaluación de tecnologías de la información. A continuación se detallan los constructos:

- **Eficacia Real:** Está conformado por dos variables de desempeño
 - **Eficiencia Real:** El esfuerzo requerido para aplicar un método. Puede ser medida utilizando variables como el tiempo requerido para aplicar el método.
 - **Efectividad Real:** El grado con el que un método alcanza sus objetivos. Puede ser medida utilizando variables como la tasa en la comprensión y el ratio de errores.
- **Eficacia Percibida:** Está conformado por dos variables de percepción.
 - **Facilidad de Uso Percibida:** El grado con el que una persona cree que un método le demandará menos esfuerzo.
 - **Utilidad Percibida:** El grado con el que una persona cree que un método alcanzará sus objetivos planteados.
- **Intención de Uso:** El grado con el que una persona está inclinada a utilizar un método.
- **Uso Real:** El grado con el que un método es utilizado en la práctica.

3.1. Planificación del Experimento

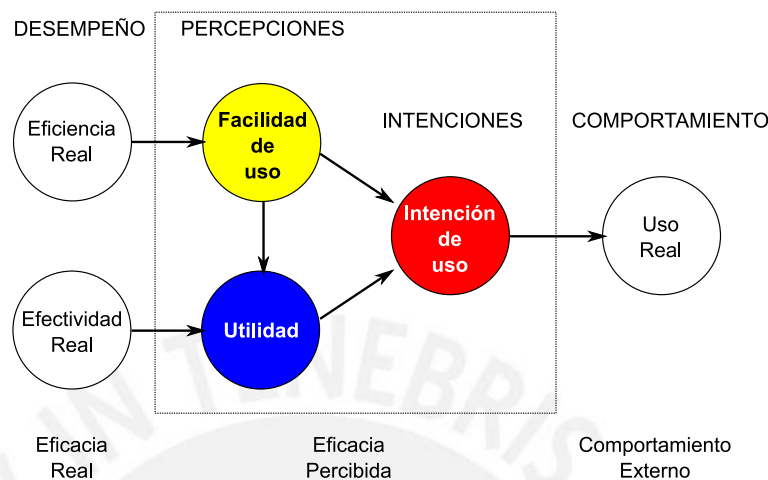


Figura 3.1: Modelo de Evaluación de Métodos propuesto por Moody et al. [18]

3.1.2. Participantes

Los participantes del experimento debieron de cumplir con 2 tareas principales: analizar los requerimientos del software a implementar, y diseñar las pantallas del sistema. A fin de evitar sesgos que pudieran presentarse durante la ejecución del experimento, se establecieron los siguientes requerimientos mínimos que los participantes debían de cumplir:

- Ser alumnos de los últimos ciclos de pregrado de la carrera de ingeniería informática ⁵.
- Contar, por lo menos, con experiencia laboral en prácticas pre-profesionales.
- Contar con experiencia en el proceso de elicitación y análisis de requerimientos de software utilizando técnicas de historias de usuario.
- Contar con experiencia en el proceso de elicitación y análisis de requerimientos de software utilizando técnicas de casos de uso.

Siendo los participantes alumnos, parcialmente se aseguraba que, en promedio, el nivel de conocimiento de todos los participantes respecto a las técnicas de elicitación fuera similar; además, al contar con una mínima experiencia laboral, les daba capacidad crítica para evaluar la utilidad del uso integrado de ambas técnicas de elicitación respecto a su uso independiente.

Quienes aceptaron participar en el experimento firmaron el Formulario de Consentimiento⁶ que les fue entregado al momento de ingresar al ambiente

⁵De acuerdo a Kitchenham et al. [55], la participación de alumnos en experimentos empíricos de ingeniería de software es aceptado debido a que ellos son la siguiente generación de profesionales.

⁶Ver Anexo B

Capítulo 3. Planteamiento del Experimento

en donde se efectuaría el estudio. Dicho documento es el único artefacto que contiene información que pueda asociar la identidad del participante a los instrumentos que serán resueltos a lo largo del estudio.

Una vez que entregaron el formulario de consentimiento, los participantes que aceptaron participar fueron agrupados de forma aleatoria en pares.

3.1.3. Instrumentos del Experimento

Para el desarrollo del experimento se hizo entrega a los participantes de los siguientes artefactos:

- **Cuestionario de Autoevaluación.** El objetivo de este artefacto es el de recabar información relacionada a la experiencia que tienen los participantes al trabajar con los modelos de casos de uso y las historias de usuario. Este documento fue completado antes de dar inicio a la experimentación (Ver Anexo C).
- **Casos Propuestos.** Este documento detalla dos casos que serán desarrollados por los participantes durante el experimento. El caso fue una historia breve que mostraba un escenario detallando las principales necesidades del interesado, su motivación y cual es el resultado que esperaba obtener una vez que estuviera implementada la solución. Sobre este documento, los participantes elaboraron los casos de uso y las historias de usuario (Ver Anexo D).
- **Registro de Historias de Usuario.** Los participantes registrarán las historias de usuario que identifiquen para cada uno de los casos propuestos (Ver Anexo E).
- **Documento de Casos de Uso.** Los participantes registraron los casos de uso y su especificación correspondiente para el caso que se les fue asignado (Ver Anexo F).
- **Cuestionario de Percepción del Análisis.** Este cuestionario fue entregado al finalizar el análisis de los casos propuestos. Estuvo conformado por un conjunto de preguntas de selección simple bajo una escala de Likert de 5 niveles [56]. Este artefacto fue utilizado para recabar información sobre la percepción de facilidad de uso, percepción de uso e intención de uso de cada uno de los métodos evaluados durante la etapa de análisis (Ver Anexo G).
- **Registro de Consultas.** Este artefacto servía como registro del número de consultas que los participantes recibieron por parte de su par. Asimismo, se indicó una breve descripción de la consulta efectuada y si fue relevante, es decir, que la información consultada no se encontraba previamente explicada en los artefactos - Historias de Usuario o Casos de Uso - o si ésta era ambigua (Ver Anexo H).

3.1. Planificación del Experimento

- **Diseño de Pantallas.** En este artefacto, los participantes dibujaron el diseño de pantallas para cada caso cuyo documento de análisis les fue sido entregado (Ver Anexo I).
- **Cuestionario de Percepción del Diseño.** Este cuestionario fue entregado al finalizar el experimento, una vez concluido el diseño de las pantallas. Estuvo conformado por un conjunto de preguntas de selección simple bajo una escala de Likert de 5 niveles [56]. Este artefacto fue utilizado para recabar información sobre la percepción de facilidad de uso, percepción de uso e intención de uso de cada uno de los métodos evaluados durante la etapa de diseño (Ver Anexo J).

3.1.4. Hipótesis, Parámetros y Variables

En base al objetivo planteado en la Sección 3.1.1, se establecieron las siguientes preguntas de investigación y sus correspondientes hipótesis nulas y alternativas.

- **Pregunta 1:** ¿Hubo un menor número de consultas realizadas al basarse en casos de uso frente a historias de usuario durante el diseño de las pantallas?
- **Pregunta 2:** ¿Los participantes, al momento de diseñar las pantallas, requieren menos tiempo para entender los casos de uso frente a las historias de usuario?
- **Pregunta 3:** ¿Los participantes experimentaron mayor facilidad para comprender los requerimientos al aplicar el método propuesto?
- **Pregunta 4:** ¿Es el método propuesto percibido como más útil?
- **Pregunta 5:** ¿Existe intención de utilizar el método propuesto en futuros proyectos?

Las hipótesis derivadas de cada una de las preguntas de investigación fueron las siguientes:

- H_{10} : El número de consultas realizadas durante el diseño de las pantallas fue el mismo al aplicar casos de uso e historias de usuario.
 H_{1a} : El número de consultas realizadas durante el diseño de las pantallas no fue el mismo al aplicar casos de uso e historias de usuario.
- H_{20} : Durante el diseño, no existe diferencia en el tiempo que se requiere para entender los casos de uso frente a las historias de usuario.
 H_{2a} : Durante el diseño, existe diferencia en el tiempo que se requiere para entender los casos de uso frente a las historias de usuario.

Capítulo 3. Planteamiento del Experimento

- H3₀: No existe diferencia en la percepción de la facilidad de uso entre el método propuesto y los métodos tradicionales.
H3_a: Existe diferencia en la percepción de la facilidad de uso entre el método propuesto y los métodos tradicionales.

- H4₀: No existe diferencia en la percepción de la utilidad entre el método propuesto y los métodos tradicionales.
H4_a: Existe diferencia en la percepción de la utilidad entre el método propuesto y los métodos tradicionales.

- H5₀: No existe intención de utilizar el método propuesto en el futuro.
H5_a: Existe intención de utilizar el método propuesto en el futuro.

Las variables independientes para cada una de las hipótesis incluyen los casos propuestos y la técnica a utilizar para la elicitación de los requerimientos que puede ser historias de usuario o su combinación con los modelos de casos de uso.

Las variables dependientes por cada una de las hipótesis planteadas se encuentran detalladas en la Tabla 3.2.

Tabla 3.2: Variables dependientes

Basada en	Nombre	Hipótesis	Métrica	Escala
Desempeño	Efectividad en la Comprensión	H1	Suma del promedio del total de consultas por técnica aplicada	Ratio
	Eficiencia en la Comprensión	H2	Suma del promedio del tiempo requerido durante el diseño para comprender cada documento de elicitación	Ratio
Percepción	Facilidad de Uso Percibida	H3	Calculado en base a las preguntas del cuestionario	
	Utilidad Percibida	H4	Calculado en base a las preguntas del cuestionario	
	Intención de Uso	H5	Calculado en base a las preguntas del cuestionario	

3.2. Diseño del Experimento

3.2.1. Procedimiento

Los participantes fueron alumnos de los últimos años de pregrado de la especialidad de ingeniería informática quienes efectuaron tareas de análisis funcional y diseño de pantallas. La investigación requirió que el participante desarrolle un breve ejercicio y responda un conjunto de cuestionarios. Todo ello tomó aproximadamente 1 hora y 40 minutos.

Los primeros 15 minutos se destinaron a la lectura de la Hoja Informativa, un documento que describe el propósito del experimento y las actividades a realizar por parte de los participantes, y a la firma del Formulario de Consentimiento por parte de los participantes que desearan colaborar de forma voluntaria con el desarrollo del experimento.

Los participantes fueron divididos de forma aleatoria en parejas. A cada uno se le hizo entrega de 1 caso conformado por 2 partes.

1. Cuestionario de Autoevaluación (5 minutos): Cada participante respondió un breve cuestionario sobre su experiencia laboral y su nivel de destreza en el manejo de técnicas para la elicitación de requerimientos; principalmente, casos de uso e historias de usuario. El objetivo de este cuestionario es el de evaluar el grado de familiaridad que tiene cada participante con las técnicas de elicitación de requerimientos y el diseño de prototipos.
2. Elicitación de requerimientos (40 minutos): Los participantes revisarán el caso que han recibido y elaborarán (a) un registro de historias de usuario y (b) un registro de historias de usuario + casos de uso. En caso consideren que falta información o ésta es ambigua, podrán consultar al evaluador.
3. Cuestionario de Análisis (5 minutos): Los participantes responderán el Cuestionario de Percepción sobre el Análisis que está conformado por 14 preguntas de selección simple. El cuestionario presentará preguntas que compararán las historias de usuario frente al método propuesto (uso conjunto de historias de usuario y casos de uso).
4. Diseño de pantallas (30 minutos): En esta sección, los participantes recibirán el registro de historias de usuario (a) y el documento de casos de uso (b) de su pareja a fin de que elaboren los prototipos de las pantallas del sistema software. Deberá de anotar la hora y minuto en el que ha recibido el documento, revisará el íntegro del documento. Si consideran que falta información o ésta es ambigua, podrá consultar a su compañero quien deberá de registrar cada consulta que le efectúen indicando si considera que la consulta ha sido relevante. Una vez que

Capítulo 3. Planteamiento del Experimento

el participante comprenda el íntegro del documento, registrará la hora y minuto. Finalmente elaborará el diseño de las pantallas en base al documento recibido.

5. Cuestionario de Diseño (5 minutos): Los participantes responderán el Cuestionario de Percepción sobre el Diseño que está conformado por 14 preguntas de selección simple. El cuestionario presentará preguntas que compararán las técnicas de historias de usuario frente a los casos de uso.

3.2.2. Validación del Experimento

Antes de ejecutar el experimento, se llevó a cabo un ejercicio para validar el mismo. Para ello se invitaron a dos profesionales expertos en el análisis, diseño e implementación de sistemas software.

Como consecuencia de la validación, se aplicaron los siguientes cambios a los instrumentos:

- Tiempo de duración del ejercicio: Se ajustaron los tiempos requeridos para las tareas de análisis y diseño.
- Formato del cuestionario de percepción: Se modificó el cuestionario para que trabaje bajo una escala de Likert de 5 niveles en lugar de 4, que fue la propuesta inicial a fin de forzar a los participantes a inclinarse por alguna de las 2 alternativas a comparar (selección forzosa propuesta por Allen et al. [57]).

Asimismo, esta actividad permitió corroborar que los 2 casos propuestos (Ver Anexo D) para el desarrollo del experimento presentaban el mismo nivel de dificultad.

3.2.3. Técnicas Estadísticas Aplicadas

Para el análisis de los resultados de cada una de las hipótesis planteadas, se determinó la necesidad de aplicar un conjunto de técnicas estadísticas.

La primera técnica aplicada sobre los datos recabados para responder cada una de las preguntas de investigación fue la prueba de bondad de ajuste. Considerando que el tamaño de la muestra era menor que 30 participantes, se optó por utilizar el método estadístico de Shapiro-Wilk [58] para determinar si los datos presentaban una distribución normal el cual, de acuerdo a Razali et al. [59], presenta un mejor desempeño que el método Kolgomorov-Smirnov.

Las hipótesis H1 y H2, buscan comparar la efectividad y la eficiencia al utilizar 2 técnicas distintas. En caso que la distribución de los datos presente una distribución normal, de acuerdo al resultado de la prueba de Shapiro-Wilk, se optará por utilizar la prueba paramétrica de comparación de medias

3.2. Diseño del Experimento

conocida como Prueba-t. Ésta permite determinar si la diferencia entre los valores promedios de dos muestras, que están siendo comparadas, es significativa [60].

En caso el resultado de la prueba de bondad de ajuste indique que la distribución de datos no seguía una distribución normal, se optará por utilizar la prueba no paramétrica de U de Mann-Whitney [61] que permitirá comparar de forma significativa dos grupos de datos en base a los valores de sus medianas.

En el caso de las hipótesis H3, H4 y H5; los datos recabados a través de las encuestas de percepción fueron comparadas con una media poblacional; para el presente estudio, considerando que la encuesta era de 5 niveles, se optó por elegir el valor 3 como media poblacional. Si la distribución de los datos presenta una distribución normal, se utilizará la Prueba-t para una muestra; caso contrario se hará uso de la prueba de Wilcoxon [62].

Finalmente, se calcularán las correlaciones entre las diversas variables analizadas a fin de determinar si para el presente estudio se cumplen las relaciones propuestas por el MEM. De acuerdo a Bonett et al. [63], en caso ambas variables a comparar presenten una distribución normal, se utilizará el coeficiente de Pearson [64], [65]; caso contrario, se analizarán los valores utilizando el coeficiente de Spearman [66].

Capítulo 3. Planteamiento del Experimento



CAPÍTULO 4

Evaluación

El objetivo de este capítulo es presentar el desarrollo del experimento previamente planificado en el capítulo anterior.

La organización de este capítulo se detalla a continuación. La Sección 4.1 describe la puesta en marcha del experimento. La Sección 4.2 presenta el análisis de resultados. La Sección 4.3 comprende la discusión de los resultados obtenidos y las amenazas a la validez del experimento. Finalmente, la Sección 4.4 presenta las conclusiones y trabajos futuros a realizar.

4.1. Ejecución del Experimento

Esta sección describe la ejecución del experimento. Se presentan las tareas llevadas a cabo como preparación para el experimento así como las desviaciones presentes que acontecieron durante la ejecución.

4.1.1. Preparación

Para llevar a cabo el experimento, se solicitó (y recibió) la aprobación por parte del Comité de Ética de Investigación con Seres Humanos y Animales (CEISHA) de la Pontificia Universidad Católica del Perú en Septiembre de 2015. El experimento fue realizado el miércoles, 7 de octubre de 2015; teniendo como participantes a los alumnos del curso de Ingeniería de Software, dictado en los últimos ciclos del pregrado en Ingeniería Informática de la PUCP. Se contó con la autorización del docente del curso previamente al desarrollo del experimento.

La evaluación se llevó a cabo dentro del horario de dictado de clases a las 15:30 - 1 hora y media luego de haber iniciada la clase.

En los primeros minutos se les asignó a cada participante un número identificador único y se les entregó a cada uno 2 documentos: La Hoja Informativa, que provee la descripción de las actividades a realizar y el tiempo asignado a cada actividad (Ver Anexo A); y el Formulario de Consentimiento, documento que recaba el consentimiento del alumno a participar en el

Capítulo 4. Evaluación

experimento (Ver Anexo B). Asimismo, el Formulario de Consentimiento es el único documento que permite relacionar la identidad real de los participantes con su número identificador.

Una vez que los Formularios de Consentimiento fueron recogidos, los participantes, quienes firmaron el documento, fueron divididos de forma aleatoria en pares.

Con ello, se dio inicio al experimento.

4.1.2. Desviaciones de la Planificación

A continuación se detallarán las circunstancias que se presentaron durante la ejecución del experimento que no estuvieron planificadas y las decisiones que se tomaron para su resolución.

La cantidad de participantes presentes en el experimento fue un número impar por lo que uno de los grupos de trabajo estuvo conformado por 3 personas. Dentro del grupo, 2 personas trabajaron el mismo caso. Se puede considerar la inclusión de un tercer caso para una futura réplica del experimento a fin de mitigar problemas similares que pudieran suscitarse.

No todos los participantes tenían experiencia en las tareas de elicitación y análisis del experimento; por ellos se aplicaron 3 cambios sobre los ejercicios que realizaron.

- Se incrementó el tiempo para la elaboración de las Historias de Usuario y los Casos de Uso en 5 minutos.
- Se decrementó en 5 minutos las tareas de Diseño de Pantallas.
- Durante el diseño de las pantallas se priorizó que los participantes entendieran los documentos elaborados en la parte de análisis frente a que concluyeran el diseño de las pantallas del sistema.

4.2. Análisis de Resultados

Las edades de los participantes estaban distribuidas entre los 19 y 40 años. Ubicándose la mayoría entre los 21 y 24 años como puede apreciarse en la Tabla 4.1.

Asimismo, la Tabla 4.2 detalla la distribución del sexo de los participantes; se contaron con 21 personas de sexo masculino y 4 de sexo femenino.

En base a los resultados del cuestionario de autoevaluación de los 25 participantes, se determinó que 7 de ellos no contaba con experiencia laboral alguna. Es por ello que la muestra se redujo a 18 personas. De los 18 participantes con experiencia laboral, solo 12 contaban con experiencia con las áreas de análisis y diseño. Sobre este conjunto de participantes se llevó a cabo análisis de datos recabados del experimento descrito en la Sección 3.2.

4.2. Análisis de Resultados

Tabla 4.1: Edad de los Participantes

Edad	Frecuencia	Porcentaje	Pct. acumulado
19	1	4.0	4.0
20	2	8.0	12.0
21	5	20.0	32.0
22	7	28.0	60.0
23	4	16.0	76.0
24	3	12.0	88.0
25	1	4.0	92.0
26	1	4.0	96.0
40	1	4.0	100.0
Total	25	100.0	

Tabla 4.2: Sexo de los Participantes

Sexo	Frecuencia	Porcentaje	Pct. acumulado
M	21	84.0	84.0
F	4	16.0	100.0
Total	25	100.0	

La Tabla 4.3 muestra el proceso de selección de participantes de acuerdo a su experiencia en las etapas de análisis y diseño del ciclo de vida del software.

Tabla 4.3: Selección de participantes de acuerdo a su experiencia en las etapas del ciclo de vida

Total de Participantes	Con Experiencia Laboral	Sólo en Análisis y Diseño
25	18	Sólo Análisis: 3
		Sólo Diseño: 0
		Ambos: 9

A continuación se procederá a analizar los datos recabados para responder cada una de las preguntas de investigación planteadas.

4.2.1. Pregunta 1: Cantidad de consultas

La pregunta 1 busca determinar si el número de consultas elaboradas durante el diseño al aplicar casos de uso frente a historias de usuario es menor. Para ello se buscará determinar la validez de las siguientes hipótesis:

- H_{10} : El número de consultas realizadas durante el diseño de las pantallas fue el mismo al aplicar casos de uso e historias de usuario.
- H_{1a} : El número de consultas realizadas durante el diseño de las pantallas no fue el mismo al aplicar casos de uso e historias de usuario.

Capítulo 4. Evaluación

Para responder esta pregunta se analizó una muestra de 12 registros de historias de usuario y 12 especificaciones de casos de uso. Se buscó determinar cuál de las 2 técnicas generó un mayor número de consultas relevantes, no relevantes y su totalizado por parte de los participantes. Las consultas no relevantes incluyeron las preguntas en las que se hacía referencia a información (posiblemente ambigua) que ya había sido detallada dentro de los documentos de casos de uso e historias de usuario. Los estadísticos descriptivos para cada uno de las técnicas aplicadas se encuentran descritos en la Tabla 4.4.

Tabla 4.4: Pregunta 1 - Estadísticos descriptivos para las consultas efectuadas durante el diseño

	Técnica aplicada	N	\bar{X}	$E_{\bar{X}}$	s	Mín	Máx
Total de Consultas	Historias de Usuario	12	1.25	0.218	0.754	0	2
	Casos de Uso	12	0.42	0.149	0.515	0	1
Consultas Relevantes	Historias de Usuario	12	0.75	0.218	0.754	0	2
	Casos de Uso	12	0.33	0.142	0.492	0	1
Consultas No Relevantes	Historias de Usuario	12	0.5	0.230	0.798	0	2
	Casos de Uso	12	0.08	0.083	0.289	0	1

Asimismo, la Figura 4.1 muestra la distribución de los datos a través de un diagrama de cajas.

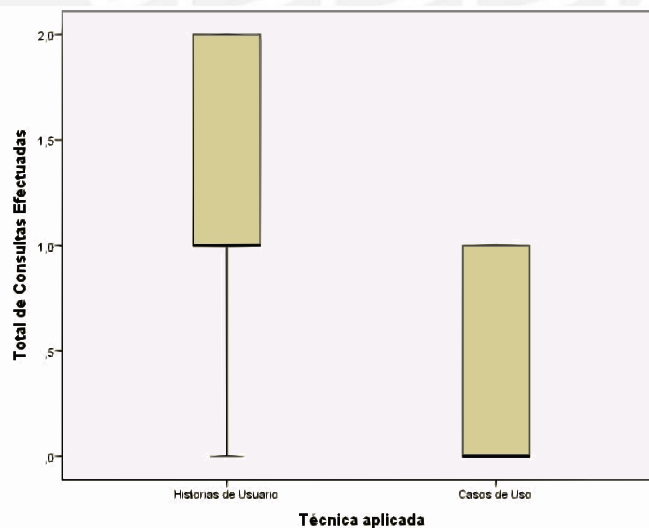


Figura 4.1: Diagrama de cajas que representan la distribución de los datos asociados al total de las consultas efectuadas durante el diseño

Debido a que el tamaño de la muestra era menor a 20 individuos, se

4.2. Análisis de Resultados

aplicó la prueba Shapiro-Wilk de bondad de ajuste para determinar si los datos presentaban una distribución normal. Los resultados en la Tabla 4.5 muestran que los datos no presentan una distribución normal con un $\alpha = 0.05$, dado que su p-valor es menor a α .

Tabla 4.5: Pregunta 1 - Resultados de la prueba de bondad de ajuste

	Técnica aplicada	Estad.	gl	Sig.
Total de Consultas	Historias de Usuario	0.81	12	0.01
	Casos de Uso	0.64	12	0.00
Consultas Relevantes	Historias de Usuario	0.81	12	0.01
	Casos de Uso	0.61	12	0.00
Consultas No Relevantes	Historias de Usuario	0.66	12	0.00
	Casos de Uso	0.33	12	0.00

Como los datos no presentaban una distribución normal, se aplicó la prueba U de Mann-Whitney para muestras independientes sobre el total de consultas, el número de consultas relevantes y el número de consultas no relevantes. Los resultados se visualizan en las Tablas 4.6 y 4.7. Con un $\alpha=0.05$ y los valores de significancia de 0.219 y 0.291 para los número de consultas relevantes y no relevantes, respectivamente, se concluye que las distribuciones son las mismas en ambos casos al utilizar cualquiera de las 2 técnicas evaluadas.

Tabla 4.6: Pregunta 1 - Valores de los rangos resultantes de la prueba de U de Mann-Whitney

	Técnica aplicada	N	Rango promedio	Suma de rangos
Total de Consultas	Historias de Usuario	12	16.04	192.50
	Casos de Uso	12	8.96	107.50
Consultas Relevantes	Historias de Usuario	12	14.33	172.00
	Casos de Uso	12	10.67	128.00
Consultas No Relevantes	Historias de Usuario	12	14.08	169.00
	Casos de Uso	12	10.92	131.00

Tabla 4.7: Pregunta 1 - Resultados de la prueba de U de Mann-Whitney

	Total de Consultas	Consultas Relevantes	Consultas No Relevantes
U de Mann-Whitney	29.500	50.000	53.000
W de Wilcoxon	107.500	128.000	131.000
Z	-2.635	-1.430	-1.548
Sig. asintótica (bilateral)	0.008	0.153	0.122
Significación exacta	0.012	0.219	0.291

Para el caso del total de consultas efectuadas, luego de aplicar la prueba U de Mann-Whitney con un $\alpha = 0.05$, se rechaza la hipótesis H_{10} con un p-

valor de 0.012. Es decir, las distribuciones del total de consultas efectuadas es distinta para cada una de las técnicas evaluadas. Se concluye que el total de consultas efectuadas al utilizar historias de usuario es mayor que al utilizar casos de uso.

4.2.2. Pregunta 2: Tiempo requerido para entender los documentos

La pregunta 2 busca determinar si los participantes requieren menos tiempo para entender los casos de uso frente a las historias de usuario al momento de diseñar las pantallas. Para ello se buscará determinar la validez de las siguientes hipótesis:

- H_{20} : Durante el diseño, no existe diferencia en el tiempo que se requiere para entender los casos de uso frente a las historias de usuario.
- H_{2a} : Durante el diseño, existe diferencia en el tiempo que se requiere para entender los casos de uso frente a las historias de usuario.

Para responder esta pregunta se analizó una muestra de 12 registros de historias de usuario y 12 especificaciones de casos de uso; sin embargo, solo fue posible utilizar 10 documentos de cada tipo debido a valores perdidos. Los estadísticos descriptivos sobre el tiempo requerido por cada participante para comprender cada documento se encuentran descritos en la Tabla 4.8. Asimismo, la Figura 4.2 muestra la distribución de los datos a través de un diagrama de cajas.

Tabla 4.8: Pregunta 2 - Estadísticos descriptivos para las consultas efectuadas durante el diseño

	Técnica aplicada	N	\bar{X}	$E_{\bar{X}}$	s	Mín	Máx
Diferencia (minutos)	Historias de Usuario	10	8.60	1.137	3.596	5	16
	Casos de Uso	10	5.80	1.236	3.910	2	13

Se aplicó la prueba Shapiro-Wilk de bondad de ajuste para determinar si los datos presentaban una distribución normal. Los resultados en la Tabla 4.9 muestran que con un $\alpha = 0.05$, los datos presentan una distribución normal con un p-valor = 0.162 para las historias de usuario y un p-valor = 0.101 para los casos de uso.

Tabla 4.9: Pregunta 2 - Resultados de la prueba de bondad de ajuste

	Técnica aplicada	Estad.	gl	Sig.
Diferencia (minutos)	Historias de Usuario	0.888	10	0.162
	Casos de Uso	0.870	10	0.101

4.2. Análisis de Resultados

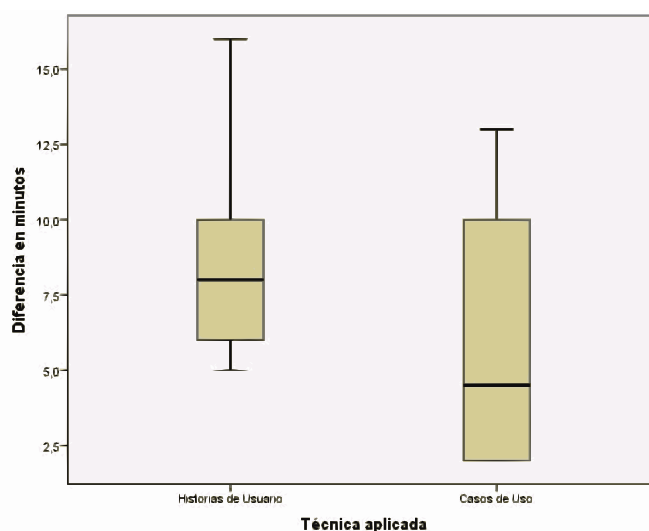


Figura 4.2: Diagrama de cajas que representan la distribución de los datos asociados al tiempo que requirieron los participantes para comprender cada uno de los documentos de requerimientos

Como los datos presentaban una distribución normal, se aplicó la prueba-t para determinar si la media del tiempo requerido para entender cada uno de los documentos elaborados durante el análisis, es diferente.

Tabla 4.10: Pregunta 2 - Resultados de la prueba de Levene

	F	Sig.
Diferencia (minutos)	0.217	0.647

De acuerdo a los datos de la Tabla 4.10, al aplicar la prueba de Levene se determinó con un $\alpha = 0,05$, que hay suficiente sustento estadístico para aceptar que las varianzas de la diferencia de tiempos al utilizar casos de uso o historias de usuario son iguales con un p-valor = 0.647.

Tabla 4.11: Pregunta 2 - Resultados de la prueba-t para 2 muestras independientes

		t	gl	Sig. (bi-lateral)	Dif. medias	Dif. err. std.	95 % intervalo de confianza
Diferencia (minutos)	Varianzas iguales	1.667	18	0.113	2.800	1.680	-0.729; 6.329
	Varianzas distintas	1.667	17.875	0.113	2.800	1.680	-0.731; 6.331

La Tabla 4.11 presenta el resultado de haber aplicado la prueba-t para la

igualdad de medias con un $\alpha = 0,05$ y una significancia de 0.113. Se concluye que no existen diferencias estadísticamente significativas entre los tiempos promedios para entender los requerimientos de las 2 muestras. Por ello, no es posible rechazar la hipótesis H_{2_0} .

4.2.3. Preguntas de Percepción (3, 4 y 5)

La pregunta 3 busca determinar si los participantes durante las etapas de análisis y diseño percibieron una mayor facilidad de uso (PEOU) al aplicar la combinación de técnicas propuestas en el estudio frente al uso independiente de las historias de usuario. Para ello se buscará determinar la validez de las siguientes hipótesis:

- H_{3_0} : No existe diferencia en la percepción de la facilidad de uso entre el método propuesto y los métodos tradicionales.
- H_{3_a} : Existe diferencia en la percepción de la facilidad de uso entre el método propuesto y los métodos tradicionales.

La pregunta 4 busca determinar si los participantes durante las etapas de análisis y diseño percibieron mayor utilidad (PU) en la combinación de técnicas propuestas en el estudio frente al uso independiente de las historias de usuario. Para ello se buscará determinar la validez de las siguientes hipótesis:

- H_{4_0} : No existe diferencia en la percepción de la utilidad entre el método propuesto y los métodos tradicionales.
- H_{4_a} : Existe diferencia en la percepción de la utilidad entre el método propuesto y los métodos tradicionales.

La pregunta 5 busca determinar si los participantes tienen intención de utilizar (ITU) la combinación de técnicas propuestas en sus futuros proyectos. Para ello se buscará determinar la validez de las siguientes hipótesis:

- H_{5_0} : No existe intención de utilizar el método propuesto en el futuro.
- H_{5_a} : Existe intención de utilizar el método propuesto en el futuro.

Previamente a analizar los datos recopilados en cada uno de los cuestionarios, fue necesario validar el constructo de los instrumentos utilizados haciendo uso de un análisis de correlación inter-item [67]. La evaluación se basó en dos criterios: Validez Convergente (CV) y Validez Divergente (DV). La primera se refiere a que la correlación entre los indicadores para medir el mismo constructo debe de ser la más alta posible; la segunda hace referencia a que la correlación entre los indicadores para medir diferentes constructos debe de ser la más baja posible. De acuerdo a Campbell et al. [67], una

4.2. Análisis de Resultados

Tabla 4.12: Preguntas 3, 4 y 5 - Validez del cuestionario de percepción de análisis

	PEOU						PU						ITU		CV	DV
	Q1	Q3	Q4	Q6	Q9	Q12	Q2	Q5	Q8	Q10	Q11	Q13	Q7	Q14		
Q1	1.00	0.69	0.49	0.30	-0.29	0.70	-0.12	-0.08	0.19	-0.25	-0.31	-0.48	0.48	0.48	0.48	-0.01
Q3	0.69	1.00	0.79	0.61	0.15	0.88	0.11	0.20	0.19	0.05	0.13	-0.38	0.67	0.76	0.69	0.22
Q4	0.49	0.79	1.00	0.62	0.13	0.83	-0.24	0.29	0.16	0.12	0.22	0.06	0.80	0.67	0.64	0.26
Q6	0.30	0.61	0.62	1.00	0.31	0.51	0.18	0.64	0.37	0.46	-0.24	-0.18	0.37	0.81	0.56	0.30
Q9	-0.29	0.15	0.13	0.31	1.00	0.32	0.30	-0.04	0.63	0.59	0.59	0.17	0.09	0.26	0.27	0.32
Q12	0.70	0.88	0.83	0.51	0.32	1.00	-0.02	-0.06	0.39	0.04	0.27	-0.24	0.73	0.67	0.71	0.22
Q2	-0.12	0.11	-0.24	0.18	0.30	-0.02	1.00	0.08	0.37	0.40	-0.08	-0.39	-0.37	0.01	0.23	-0.02
Q5	-0.08	0.20	0.29	0.64	-0.04	-0.06	0.08	1.00	0.12	0.55	-0.20	0.31	0.03	0.50	0.31	0.18
Q8	0.19	0.19	0.16	0.37	0.63	0.39	0.37	0.12	1.00	0.72	0.19	0.21	-0.09	0.31	0.43	0.27
Q10	-0.25	0.05	0.12	0.46	0.59	0.04	0.40	0.55	0.72	1.00	0.23	0.42	-0.07	0.34	0.55	0.16
Q11	-0.31	0.13	0.22	-0.24	0.59	0.27	-0.08	-0.20	0.19	0.23	1.00	0.50	0.27	0.05	0.28	0.12
Q13	-0.48	-0.38	0.06	-0.18	0.17	-0.24	-0.39	0.31	0.21	0.42	0.50	1.00	-0.12	-0.15	0.34	-0.16
Q7	0.48	0.67	0.80	0.37	0.09	0.73	-0.37	0.03	-0.09	-0.07	0.27	-0.12	1.00	0.66	0.83	0.23
Q14	0.48	0.76	0.67	0.81	0.26	0.67	0.01	0.50	0.31	0.34	0.05	-0.15	0.66	1.00	0.83	0.39

pregunta dentro del cuestionario se considerará como válida si el valor de su CV es mayor que su DV.

En la Tabla 4.12 se puede apreciar que la pregunta Q9 tiene un valor de DV (0.32) mayor que el CV (0.27). Por esta razón, Q9 fue excluido del análisis de datos.

Tabla 4.13: Preguntas 3, 4 y 5 - Validez del cuestionario de percepción de diseño

	PEOU						PU						ITU		CV	DV
	Q3	Q4	Q6	Q8	Q9	Q12	Q1	Q2	Q5	Q10	Q11	Q13	Q7	Q14		
Q3	1.00	-0.07	0.23	0.28	-0.31	0.41	0.06	0.59	0.14	0.23	0.09	0.28	-0.02	-0.02	0.26	0.17
Q4	-0.07	1.00	0.34	0.13	-0.63	0.36	0.73	0.52	0.43	0.30	0.95	0.13	0.51	0.51	0.19	0.51
Q6	0.23	0.34	1.00	0.33	0.10	0.47	0.33	0.30	0.57	-0.23	0.25	0.33	0.67	0.67	0.41	0.36
Q8	0.28	0.13	0.33	1.00	0.22	0.91	0.73	0.70	0.89	0.50	0.31	1.00	0.73	0.73	0.48	0.70
Q9	-0.31	-0.63	0.10	0.22	1.00	-0.10	-0.23	-0.28	0.03	-0.29	-0.60	0.22	-0.03	-0.03	0.05	-0.15
Q12	0.41	0.36	0.47	0.91	-0.10	1.00	0.82	0.74	0.94	0.61	0.47	0.91	0.85	0.85	0.51	0.77
Q1	0.06	0.73	0.33	0.73	-0.23	0.82	1.00	0.78	0.84	0.68	0.82	0.73	0.77	0.77	0.81	0.50
Q2	0.59	0.52	0.30	0.70	-0.28	0.74	0.78	1.00	0.63	0.57	0.73	0.70	0.45	0.45	0.73	0.43
Q5	0.14	0.43	0.57	0.89	0.03	0.94	0.84	0.63	1.00	0.44	0.49	0.89	0.96	0.96	0.72	0.61
Q10	0.23	0.30	-0.23	0.50	-0.29	0.61	0.68	0.57	0.44	1.00	0.45	0.50	0.31	0.31	0.61	0.22
Q11	0.09	0.95	0.25	0.31	-0.60	0.47	0.82	0.73	0.49	0.45	1.00	0.31	0.48	0.48	0.63	0.30
Q13	0.28	0.13	0.33	1.00	0.22	0.91	0.73	0.70	0.89	0.50	0.31	1.00	0.73	0.73	0.69	0.54
Q7	-0.02	0.51	0.67	0.73	-0.03	0.85	0.77	0.45	0.96	0.31	0.48	0.73	1.00	1.00	1.00	0.53
Q14	-0.02	0.51	0.67	0.73	-0.03	0.85	0.77	0.45	0.96	0.31	0.48	0.73	1.00	1.00	1.00	0.53

En la Tabla 4.13 se puede apreciar que las preguntas Q4, Q8 y Q12 tiene un valor de DV mayor que el CV. Por esta razón, las preguntas en mención fueron excluidas del análisis de datos.

Asimismo, según Straub [68], el uso de múltiples preguntas para medir el mismo constructo requiere la validación de su consistencia interna; es decir, validar que cada vez que el instrumento de medición sea utilizado, éste debe de retornar resultados similares. Para validar su confiabilidad, se hizo uso del α de Chronbach. El valor del α de Chronbach para cada uno de los cuestionarios, luego de remover las preguntas Q9 (para análisis); y Q4, Q8 y Q12 (para diseño) fue de 0.795 y 0.36 tal como se presenta en la Tabla 4.14. Siendo 0.7, el mínimo valor recomendado por Nunnally [69] para estudios exploratorios.

Para proceder con el análisis de los datos, se procedió a promediar por

Tabla 4.14: Pregunta 3, 4 y 5 - Confiabilidad inicial de los cuestionarios de percepción

		Chronbach (α)	N
Análisis	PEOU	0.827	5
	PU	0.750	6
	ITU	0.807	2
Total		0.795	13
Diseño	PEOU	-0.634	3
	PU	0.857	6
	ITU	0.859	2
Total		0.36	11

participante, cada una de las respuestas que busquen evaluar la misma categoría de percepción - PEOU, PU, ITU. La Tabla 4.15 muestra los estadísticos descriptivos para las respuestas brindadas durante cada etapa.

Tabla 4.15: Preguntas 3, 4 y 5 - Estadísticos descriptivos para los cuestionarios de percepción de análisis y diseño

		PEOU	PU	ITU
Análisis	N	12	12	12
	\bar{X}	3.6333	3.5983	3.3750
	$E_{\bar{X}}$	0.23720	0.16452	0.25469
	s	0.82168	0.56993	0.88227
	Mín.	2.20	2.67	2.00
	Máx.	4.80	4.67	5.00
Diseño	N	8	8	8
	\bar{X}	3.3125	3.3125	3.5625
	$E_{\bar{X}}$	0.18537	0.35481	0.41659
	s	0.52432	1.00356	1.17830
	Mín.	2.33	1.33	1.00
	Máx.	4.00	4.17	5.00

Sobre los datos, se aplicó la prueba Shapiro-Wilk de bondad de ajuste para determinar si los datos presentaban una distribución normal. Los resultados en la Tabla tab:evaluacion-percepcion-normalidad muestran:

- Análisis: Los datos asociados a PEOU, PU, ITU presentan una distribución normal con un $\alpha = 0,05$ y niveles de significancia de 0.052, 0.967 y 0.393, respectivamente, que son mayores a α .
- Diseño: Los datos asociados a PEOU, PU presentan una distribución normal con un $\alpha = 0,05$ debido a que sus niveles de significancia de 0.672 y 0.066 son mayores que α . Sin embargo, los datos asociados a ITU no presentan una distribución normal con una significancia de 0.039 que es menor a $\alpha = 0,05$.

4.2. Análisis de Resultados

Tabla 4.16: Preguntas 3, 4 y 5 - Resultados de la prueba de bondad de ajuste

		Estad.	gl	Sig.
Análisis	PEOU	0.862	12	0.052
	PU	0.977	12	0.967
	ITU	0.931	12	0.393
Diseño	PEOU	0.946	8	0.672
	PU	0.834	8	0.066
	ITU	0.821	8	0.047

En base a los resultados en la Tabla 4.17 de la prueba-t sobre los valores recopilados vía el cuestionario de percepción de análisis, se concluye con un $\alpha = 0,05$, que para la facilidad de uso percibida ($p = 0.022$) y para la utilidad percibida ($p = 0.004$) se rechaza la hipótesis nula. Es decir, existe una diferencia estadística significativa entre el valor promedio de la percepción de facilidad/utilidad de la muestra y el promedio de la población ($\mu \neq 3$). Sin embargo, no existe una diferencia estadística significativa entre el valor medio de la intención de uso de la muestra y el promedio de la población con un $p = 0.169$ ($\mu = 3$). Se concluye que los valores promedios de la muestra para la percepción de facilidad ($\bar{X} = 3,6333$) y de utilidad ($\bar{X} = 3,5983$) es estadísticamente alta ($\mu < \bar{X}$).

Tabla 4.17: Preguntas 3, 4 y 5 - Resultados de la prueba-t para 1 muestra con un $\mu = 3$ sobre los datos recabados a través del cuestionario de percepción del análisis

	t	gl	Sig. (bilateral)	Dif. medias	95 % intervalo de confianza
PEOU	2.670	11	0.022	0.63333	0.1113 / 1.1554
PU	3.637	11	0.004	0.59833	0.2362 / 0.9604
ITU	1.472	11	0.169	0.37500	-0.1856 / 0.9356

En base a los resultados en la Tabla 4.18 de la prueba-t sobre los valores recopilados vía el cuestionario de percepción de diseño, se llevó a cabo una prueba-t sobre las variables PEOU y PU. Se concluye con un $\alpha = 0,05$, que para la facilidad de uso percibida ($p = 0.136$) y la utilidad percibida ($p = 0.408$) no es posible rechazar la hipótesis nula. Es decir, no existe una diferencia estadística significativa entre el valor promedio de la percepción de facilidad/utilidad de la muestra y el promedio de la población ($\mu = 3$).

Para analizar los datos sobre la intención de uso de la técnica propuesta durante la etapa de diseño, se efectuó la prueba de Wilcoxon sobre la variable ITU cuyos resultados se visualizan en la Tabla 4.19. Con una significancia de 0.196 y un $\alpha = 0,05$ para una mediana hipotética de valor 3; no es posible rechazar la hipótesis nula. Por ello, no es posible concluir que la mediana de la intención de uso de la muestra sea mayor a la mediana promedio hipotética para la población.

Capítulo 4. Evaluación

Tabla 4.18: Preguntas 3 y 4 - Resultados de la prueba-t para 1 muestra con un $\mu = 3$ sobre los datos recabados a través del cuestionario de percepción del diseño

	t	gl	Sig. (bilateral)	Dif. medias	95 % intervalo de confianza
PEOU	1.686	7	0.136	0.31250	-0.1258 / 0.7508
PU	0.881	7	0.408	0.31250	-0.5265 / 0.1515

Tabla 4.19: Pregunta 5 - Resultados de la prueba de Wilcoxon para 1 muestra con una mediana = 3 sobre los datos recabados a través del cuestionario de percepción del diseño

N	Estad. Contraste	Error Std.	Estad. Contraste Std.	Sig. (bilateral)
8	21.500	5.799	1.293	0.196

Finalmente, para la etapa de análisis se aceptan las hipótesis $H3_a$ y $H4_a$; sin embargo, no hay suficiente evidencia para rechazar la hipótesis $H5_0$. En la etapa de diseño, no es posible rechazar ninguna de las hipótesis nulas $H3_0$, $H4_0$ y $H5_0$.

4.2.4. Análisis de las Relaciones del MEM

El objetivo de esta sección es evaluar las relaciones del modelo MEM propuestas por Moody et al. [18] y que se muestran en la Figura 3.1.

Para el análisis de las fortalezas de las correlaciones entre las variables estudiadas, Mujis [70] propone un conjunto de reglas agrupadas de acuerdo al rango del coeficiente de correlación. Las reglas se muestran en la Tabla 4.20.

Tabla 4.20: Correlaciones - Reglas propuestas por Mujis para determinar la fortaleza de la relación entre 2 variables

Rango del valor absoluto del índice de correlación	Fortaleza de la relación
[0,0,0,1]	Débil
[0,1,0,3]	Modesto
[0,3,0,5]	Moderado
[0,5,0,8]	Fuerte
[0,8,1,0]	Muy Fuerte

Correlación de datos durante el análisis

Se procedió con el análisis de las correlaciones entre los datos recopilados de participantes que cuentan con experiencia en la etapa de análisis del ciclo de vida del software. Las Tablas 4.21 y 4.22 presentan el resumen y los

4.2. Análisis de Resultados

descriptivos de los datos sobre los que se realizó el análisis. De ellos, el 25.0% son valores perdidos.

Tabla 4.21: Correlaciones - Resumen de los datos recopilados de participantes con experiencia en análisis

	Válido		Perdidos		Total	
	N	Pct	N	Pct	N	Pct
Delta de Tiempo	9	75.0 %	3	25.0 %	12	100.0 %
Delta de Consultas	9	75.0 %	3	25.0 %	12	100.0 %
PEOU	9	75.0 %	3	25.0 %	12	100.0 %
PU	9	75.0 %	3	25.0 %	12	100.0 %
ITU	9	75.0 %	3	25.0 %	12	100.0 %

Tabla 4.22: Correlaciones - Descriptivos de los datos recopilados de participantes con experiencia en análisis

	Delta Tiempo	Delta Consultas	PEOU	PU	ITU
\bar{X}	-2.78	-0.78	3.667	3.447	3.389
$E_{\bar{X}}$	1.839	0.324	0.302	0.191	0.320
s	5.518	0.972	0.906	0.572	0.961
Mín.	-10	-2	2.20	2.67	2.00
Máx.	7	0	4.80	4.67	5.00

De acuerdo a la Tabla 4.23, luego de ejecutar la prueba de Shapiro-Wilk de bondad de ajuste con un $\alpha = 0,05$, se tiene que el delta de tiempo, el PU y el ITU presentan una distribución normal con una significancia de 0.195, 0.303 y 0.654 respectivamente. El delta de consultas y el PEOU, presentan una distribución no normal con una significancia de 0.002 y 0.048 respectivamente.

Tabla 4.23: Correlaciones - Resultados de la prueba de bondad de ajuste de los datos recopilados de participantes con experiencia en análisis

	Estad.	gl	Sig.
Delta de Tiempo	0.889	9	0.195
Delta de Consultas	0.711	9	0.002
PEOU	0.833	9	0.048
PU	0.908	9	0.303
ITU	0.947	9	0.654

Como se puede apreciar en la Tabla 4.24, no ha sido posible validar todas las relaciones entre todas las variables presentes en el modelo MEM durante el análisis. En base a la muestra, se ha determinado que existe una relación modesta entre el Delta del Tiempo vs. la Percepción de Facilidad de Uso y una relación moderada entre la Intención de Uso vs. la Percepción de Facilidad de Uso y la Intención de Uso vs. la Percepción de Utilidad; sin embargo, no es posible generalizar estos resultados con un $\alpha = 0,05$.

Capítulo 4. Evaluación

Tabla 4.24: Correlaciones - Resumen de correlaciones de los datos recopilados de participantes con experiencia en análisis

Variable 1	Variable 2	r	ρ	p-valor	Fortaleza
Delta de Tiempo	PEOU		-0.246	0.384	Modesto
Delta de Consultas	PU		-0.019	0.939	Débil
PEOU	PU		-0.067	0.776	Débil
PEOU	ITU		-0.474	0.050	Moderado
PU	ITU	0.462		0.130	Moderado

Correlación de datos durante el diseño

Se procedió con el análisis de las correlaciones entre los datos recopilados de participantes que cuentan con experiencia en la etapa de diseño del ciclo de vida del software. Las Tablas 4.25 y 4.26 presentan el resumen y los descriptivos de los datos sobre los que se realizó el análisis. De ellos, el 33.0 % son valores perdidos.

Tabla 4.25: Correlaciones - Resumen de los datos recopilados de participantes con experiencia en diseño

	Válido		Perdidos		Total	
	N	Pct	N	Pct	N	Pct
Delta de Tiempo	6	66.7 %	3	33.3 %	9	100.0 %
Delta de Consultas	6	66.7 %	3	33.3 %	9	100.0 %
PEOU	6	66.7 %	3	33.3 %	9	100.0 %
PU	6	66.7 %	3	33.3 %	9	100.0 %
ITU	6	66.7 %	3	33.3 %	9	100.0 %

Tabla 4.26: Correlaciones - Descriptivos de los datos recopilados de participantes con experiencia en diseño

	Delta Tiempo	Delta Consultas	PEOU	PU	ITU
\bar{X}	-3.00	-0.67	3.3617	3.083	3.25
$E_{\bar{X}}$	1.77	0.422	0.125	0.437	0.478
s	4.336	1.033	0.307	1.07	1.173
Mín.	-8	-2	3	1.33	1.00
Máx.	5	0	3.67	4.17	4.00

De acuerdo a la Tabla 4.27, luego de ejecutar la prueba de Shapiro-Wilk de bondad de ajuste con un $\alpha = 0,05$, se tiene que el delta de tiempo, el PEOU y el PU presentan una distribución normal con una significancia de 0.201, 0.162 y 0.475 respectivamente. El delta de consultas y el ITU, presentan una distribución no normal con una significancia de 0.001 y 0.015 respectivamente.

Como se puede apreciar en la Tabla 4.28, ha sido posible validar todas las relaciones entre todas las variables presentes en el modelo MEM durante

4.3. Discusión

Tabla 4.27: Correlaciones - Resultados de la prueba de bondad de ajuste de los datos recopilados de participantes con experiencia en diseño

	Estad.	gl	Sig.
Delta de Tiempo	0.863	6	0.201
Delta de Consultas	0.640	6	0.001
PEOU	0.851	6	0.162
PU	0.916	6	0.476
ITU	0.739	6	0.015

el diseño. En base a la muestra, se ha determinado que existe una relación modesta o moderada modesta entre las variables analizadas. Sin embargo, no es posible generalizar estos resultados con un $\alpha = 0,05$.

Tabla 4.28: Correlaciones - Resumen de correlaciones de los datos recopilados de participantes con experiencia en diseño

Variable 1	Variable 2	r	ρ	p-valor	Fortaleza
Delta de Tiempo	PEOU	0.252		0.630	Modesto
Delta de Consultas	PU		-0.354	0.291	Moderado
PEOU	PU	0.233		0.579	Modesto
PEOU	ITU		0.460	0.138	Moderado
PU	ITU		0.376	0.225	Moderado

4.3. Discusión

4.3.1. Evaluación de Resultados e Implicancias

A continuación se procederá a responder cada una de las preguntas de investigación en base a la información recabada durante el análisis de datos en la Sección 4.2.

Pregunta 1

¿Hubo un menor número de consultas efectuadas al basarse en casos de uso frente a historias de usuario durante el diseño de las pantallas? En base a los datos recabados y analizados en la Sección 4.2.1, los resultados obtenidos indican de forma significativa que al utilizar casos de uso, hubieron un menor número de consultas que cuando se utilizaron historias de usuario. Sin embargo, no es posible establecer una conclusión respecto a si hubieron un mayor número de consultas relevantes o no relevantes.

Pregunta 2

¿Los participantes, al momento de diseñar las pantallas, requieren menos tiempo para entender los casos de uso frente a las historias de usuario?

Capítulo 4. Evaluación

En base a los datos recabados y analizados en la Sección 4.2.2, se puede visualizar que el tiempo requerido para comprender las historias de usuario fue mayor que el tiempo para comprender los casos de uso con una diferencia de 2.8 minutos en promedio; sin embargo, estadísticamente los resultados obtenidos no son significativamente concluyentes por lo que no es posible establecer que el tiempo requerido para comprender los casos de usos es menor que para las historias de usuario.

Pregunta 3

¿Los participantes experimentaron mayor facilidad para comprender los requerimientos al aplicar el método propuesto? En base a los datos recabados y analizados en la Sección 4.2.3, los resultados obtenidos indican que en promedio durante la etapa de análisis los participantes percibieron que al utilizar el método propuesto les fue más fácil comprender los requerimientos con un valor de 3.63 / 5.

Durante la etapa de diseño, se obtuvo un índice de facilidad de uso de 3.31 / 5; sin embargo, no es posible generalizar este resultado debido a que estadísticamente no es significativo.

Pregunta 4

¿Es el método propuesto percibido como más útil? En base a los datos recabados y analizados en la Sección 4.2.3, los resultados obtenidos indican que en promedio durante la etapa de análisis los participantes percibieron que el método propuesto les pareció más útil con un valor de 3.59 / 5.

Durante la etapa de diseño, se obtuvo un índice de facilidad de uso de 3.31 / 5; sin embargo, no es posible generalizar este resultado debido a que estadísticamente no es significativo.

Pregunta 5

¿Existe intención de utilizar el método propuesto en futuros proyectos? En base a los datos recabados y analizados en la Sección 4.2.3, los resultados obtenidos indican que en promedio durante la etapas de análisis los participantes indicaron que en promedio optarían por utilizar el método propuesto con un valor de 3.37 / 5.

Durante la etapa de diseño, se obtuvo una mediana de intención de uso de 4 / 5.

Sin embargo, en ambos casos no es posible generalizar este resultado debido a que estadísticamente no son significativos.

4.3.2. Amenazas a la Validez

De acuerdo a Jedlitschka et al. [53] todas aquellas amenazas que puedan impactar en los resultados obtenidos han de ser discutidos. Se identifican 4 posibles amenazas a la validez.

Validez del Constructo

Para la recolección de los datos se hicieron uso de 2 casos de estudio que tenían un nivel de dificultad de resolución similar a fin de evitar un sesgo que pudiera generarse por conocimiento ya aprendido. Asimismo, se contó con un cuestionario por cada rol asignado que contenía 14 preguntas de percepción que permitían una calificación en base a una escala de Likert de 5 niveles. El orden de los elementos fue aleatorio y la mitad de las preguntas fueron formuladas de forma negativa para evitar las respuestas monótonas. Sin embargo, no ha sido posible garantizar el nivel de confiabilidad del instrumento utilizado para evaluar la facilidad de uso del método propuesto durante la etapa de diseño que ha presentado un valor negativo del α de Chronbach.

Validez Interna

A fin de evitar que se produjera un sesgo por el cansancio de los participantes en el estudio, se buscó que el experimento no tuviera una duración mayor a las 2 horas. Asimismo, se buscó que los participantes fueran alumnos pertenecientes al mismo curso de la carrera de Ingeniería Informática y que hubieran realizado prácticas profesionales con el objetivo de que los participantes no tuvieran un nivel de conocimiento y experiencia distintos; sin embargo, este objetivo no pudo ser alcanzado debido a que los participantes presentaron un nivel de experiencia profesional muy variable.

Validez Externa

En base a los resultados obtenidos no es posible asegurar estadísticamente una validez externa de este estudio para todas las preguntas de investigación planteadas. Se recomienda efectuar réplicas del experimento con un mayor número de individuos que cuenten con mayor experiencia laboral en el campo a fin de obtener resultados concluyentes.

Validez de las Conclusiones

Se hicieron uso de técnicas estadísticas a fin de mitigar el riesgo de obtener conclusiones incorrectas por una interpretación errónea de los datos.

4.4. Conclusiones de la Evaluación

Luego de haber efectuado el experimento y de haber analizado y discutido los datos recabados se puede concluir que los participantes requirieron efectuar un menor número de consultas cuando trabajaron con casos de uso. Esto puede deberse a la mayor cantidad de información que la especificación de casos de uso provee frente a las historias de usuario, lo cual tiende a introducir una menor ambigüedad al comprender los requerimientos ya recabados. Asimismo, para los participantes que participaron en el experimento, les toma menos tiempo comprender los requerimientos especificados en casos de uso; mientras que el analizar las historias de usuario recabadas les demanda un mayor tiempo. Sin embargo, para poder generalizar este resultado, es necesario trabajar con una muestra de mayor tamaño.

Durante la etapa de análisis, los participantes percibieron que el uso de casos de uso en conjunto con las historias de usuario les resultaba más fácil y le encontraban una mayor utilidad frente a utilizar las historias de usuario de forma independiente. Estos resultados fueron similares a los obtenidos durante la etapa de diseño; sin embargo, no pueden generalizarse debido a la poca experiencia con la que contaban los participantes al trabajar con casos de uso.

A pesar de que los valores relacionados a la intención de uso de ambas técnicas en conjunto durante el análisis y el diseño no arrojan resultados concluyentes. Se puede afirmar que dentro de la muestra estudiada, los participantes tienen una intención de utilizar la combinación de ambas técnicas en futuros proyectos.

Finalmente, el valor del α de Chronbach sobre la confiabilidad de las preguntas de percepción en facilidad de uso durante el diseño puede deberse a la confusión generada por la redacción de las preguntas; o que durante la elaboración de las preguntas se asumió que los participantes presentaban capacidades con las que realmente no contaban.

A fin de mitigar las amenazas en la validez detectadas en el presente estudio, se recomienda efectuar una réplica del experimento sobre una muestra de mayor tamaño en la que los participantes cuenten con una mayor experiencia laboral. Asimismo, reducir el número de preguntas de percepción a 3 por cada variable evaluada para evitar confundir a los participantes con la complejidad de las preguntas.

CAPÍTULO 5

Conclusiones

La elicitación de requerimientos es una de las principales tareas que debe llevarse a cabo para la correcta implementación de un desarrollo software. Su incorrecta especificación genera costos innecesarios a lo largo del proyecto e inclusive, su completo fracaso. Actualmente existen diversas técnicas que sirven como apoyo a la elicitación de requerimientos; siendo las historias de usuario y el modelo de casos de uso 2 de las más conocidas.

5.1. Revisión Sistemática

El presente estudio, inició con el desarrollo de una revisión sistemática que tenía como objetivo la evaluación de modelos que combinan los modelos de casos de uso y las historias de usuario. Se elaboraron 3 preguntas de investigación que fueron respondidas en base a los datos recabados de los 11 artículos relevantes identificados en el proceso de búsqueda.

A la fecha de la revisión, existen 3 modelos propuestos - Athena, K-gileRE y NORMAP - que combinan las historias de usuario con los casos de uso. Las 3 propuestas han sido validadas en entornos controlados; sin embargo, como ninguna ha sido evaluada en un entorno no controlado, no es posible determinar su eficacia en el ámbito industrial.

Además, los 3 métodos presentan características y asunciones que no necesariamente aplican en un entorno de trabajo real. En el caso de Athena y K-gileRE, los autores asumen que los interesados contarán con la disponibilidad y el interés de trabajar de forma colectiva entre ellos y con los analistas; así como la predisposición de los analistas a compartir el conocimiento - del análisis de los requisitos - con los interesados. No existen garantías que aseguren el cumplimiento de esas asunciones debido a que no se puede afirmar que las personas participantes en el proyecto compartan una misma ideología sobre como llevar a cabo su trabajo.

En el caso del método NORMAP, el modelo ha sido pensado para su aplicación en proyectos que hacen uso de metodologías ágiles; mas no hay información y validaciones efectuadas sobre su eficiencia en proyectos que

no trabajan con dicha metodología. Cabe mencionar que a pesar de que NORMAP ha sido construido para ser aplicado en proyectos ágiles, la alta complejidad del mismo genera una fuerte dependencia con su herramienta CASE (NORMATIC).

Asimismo, se encontraron pocos estudios que evaluaban la efectividad del uso combinado de las historias de usuario y los casos de uso. Esto motivó el desarrollo de la segunda parte de este estudio.

5.2. Evaluación Empírica

La segunda parte del presente trabajo fue la planificación y ejecución de un experimento empírico en el que participaron alumnos de últimos ciclos del pregrado de ingeniería informática.

Del análisis de los datos derivados del experimento se puede concluir que los participantes requirieron efectuar un menor número de consultas cuando trabajaron con casos de uso. Esto puede deberse a la mayor cantidad de información que la especificación de casos de uso provee frente a las historias de usuario, lo cual tiende a introducir una menor ambigüedad al comprender los requerimientos ya recabados.

Respecto al tiempo que demandó a los participantes en el experimento el comprender los requerimientos especificados, al trabajar con casos de uso les tomó un menor tiempo en comparación con las historias de usuario. Sin embargo, para poder generalizar este resultado, es necesario trabajar con una muestra de mayor tamaño.

Al analizar la percepción de los participantes al utilizar ambas técnicas, se obtuvo que durante la etapa de análisis el uso en conjunto de las historias de usuario con los casos de uso fue percibido como más fácil y más útil frente a utilizar las historias de usuario de forma independiente. Estos resultados fueron similares a los obtenidos durante la etapa de diseño; sin embargo, no pueden generalizarse debido a la poca experiencia con la que contaban los participantes al trabajar con casos de uso.

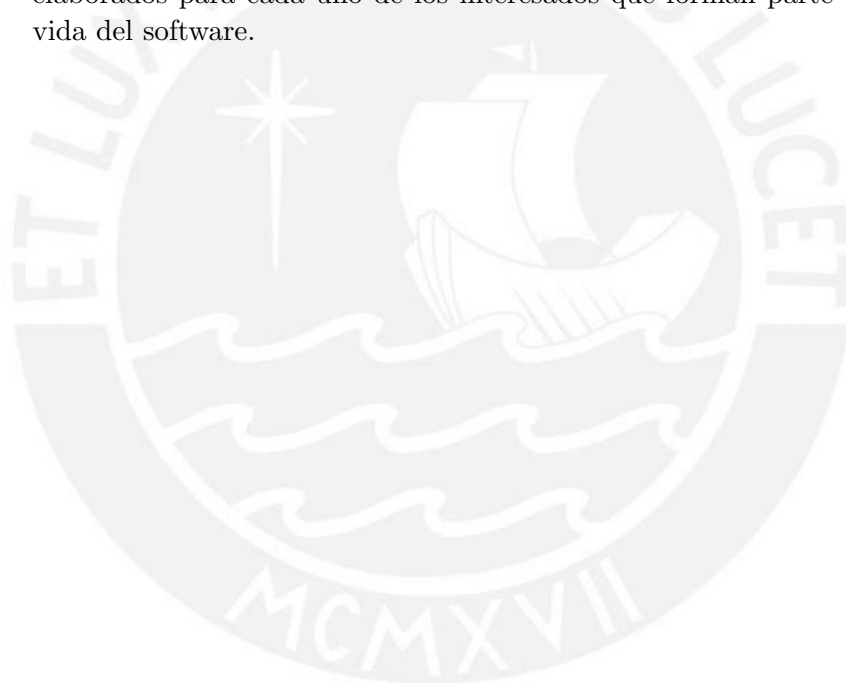
A pesar de que los valores relacionados a la intención de uso de ambas técnicas en conjunto durante el análisis y el diseño no arrojan resultados concluyentes. Se puede afirmar que dentro de la muestra estudiada, los participantes tienen una intención de aplicar ambas técnicas en conjunto en futuros proyectos.

Cabe notar el bajo valor del α de Chronbach sobre la confiabilidad en las preguntas de percepción en facilidad de uso durante el diseño. Esto puede deberse a la confusión generada por la redacción de las preguntas; o que durante la elaboración de las preguntas se asumió que los participantes presentaban capacidades con las que realmente no contaban.

5.3. Trabajo Futuro

A fin de mitigar las amenazas en la validez detectadas en el presente estudio, se propone efectuar una réplica del experimento sobre una muestra de mayor tamaño en la que los participantes cuenten con una mayor experiencia en el ciclo de vida del desarrollo de software. Asimismo, reducir el número de preguntas de percepción a 3 por cada variable evaluada para evitar confundir a los participantes con la complejidad de las preguntas.

Finalmente, es posible extender el alcance de este estudio incluyendo una evaluación de la conformidad de los requerimientos iniciales contra los diseños elaborados en base a cada una de las técnicas a aplicadas. Además de evaluar la transparencia de los casos de uso y las historias de usuario elaborados para cada uno de los interesados que forman parte del ciclo de vida del software.



Capítulo 5. Conclusiones



Bibliografía

- [1] B. W. Boehm, R. K. McClean, and D. B. Urfrig, “Some experience with automated aids to the design of large-scale reliable software,” *ACM SIGPLAN Notices*, vol. 10, no. 6, pp. 105–113, 1975.
- [2] Standish Group, “CHAOS Report,” tech. rep., 1995.
- [3] M. R. Barbacci, R. Ellison, A. J. Lattanze, J. a. Stafford, C. B. Weinstock, and W. G. Wood, “Quality Attribute Workshops, Third Edition,” Tech. Rep. August, 2003.
- [4] R. L. Nord and J. E. Tomayko, “Software architecture-centric methods and agile development,” *IEEE Software*, vol. 23, no. 2, pp. 47–53, 2006.
- [5] K. Bittner, *Use Case Modeling*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2002.
- [6] A. Abran, J. W. Moore, P. Bourque, R. Dupuis, and L. Tripp, *Guide to the Software Engineering Body of Knowledge*. IEEE, 3 ed., 2004.
- [7] ISO/IEC JTC 1, “ISO/IEC 25010:2011,” 2011.
- [8] I. Alexander and T. Zink, “Introduction to systems engineering with use cases,” *Computing and Control Engineering Journal*, vol. 13, no. 6, pp. 289–297, 2002.
- [9] R. E. Gallardo-Valencia, V. Olivera, and S. E. Sim, “Are Use Cases Beneficial for Developers Using Agile Requirements?,” in *2007 Fifth International Workshop on Comparative Evaluation in Requirements Engineering*, pp. 11–22, IEEE, 2007.
- [10] A. Cockburn, *Writing effective use cases*. 2001.
- [11] M. Cohn, *User Stories Applied: For Agile Software Development*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 2004.
- [12] K. Winbladh, H. Ziv, and D. J. Richardson, “Surveying the usability of requirements approaches using a 3-dimensional framework,” tech. rep., University of California, 2008.

Bibliografía

- [13] M. Hvalshagen, V. Khatri, and R. Venkataraman, “Understanding use cases: Harnessing the power of narratives to comprehend application domains,” in *14th Americas Conference on Information Systems, AMCIS 2008*, vol. 5, pp. 3328–3335, 2008.
- [14] Y.-C. Tu, C. Thomborson, and E. Tempero, *Transparency in Software Engineering*. Phd thesis, The University of Auckland, 2014.
- [15] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. Project Management Institute, Incorporated, 2013.
- [16] S. España, N. Condori-Fernandez, A. González, and Ó. Pastor, “Evaluating the completeness and granularity of functional requirements specifications: A controlled experiment,” in *Proceedings of the IEEE International Conference on Requirements Engineering*, pp. 161–170, 2009.
- [17] S. Abrahão, E. Insfran, J. A. Carsí, and M. Genero, “Evaluating requirements modeling methods based on user perceptions: A family of experiments,” *Information Sciences*, vol. 181, no. 16, pp. 3356–3378, 2011.
- [18] D. L. Moody, G. Sindre, T. Brasethvik, and A. Solvberg, “Evaluating the Quality of Process Models: Empirical Analysis of a Quality Framework,” *21st International Conference on Conceptual Modeling (ER 2002)*, pp. 380–396, 2002.
- [19] F. Wanderley and D. S. Da Silveria, “A framework to diminish the gap between the business specialist and the software designer,” *Proceedings - 2012 8th International Conference on the Quality of Information and Communications Technology, QUATIC 2012*, no. Figure 1, pp. 199–204, 2012.
- [20] F. Wanderley, D. Silveira, J. Araujo, A. Moreira, and E. Guerra, “Experimental Evaluation of Conceptual Modelling through Mind Maps and Model Driven Engineering,” pp. 200–214, 2014.
- [21] I. Hadar, I. Reinhartz-Berger, T. Kuflik, A. Perini, F. Ricca, and A. Susi, “Comparing the comprehensibility of requirements models expressed in Use Case and Tropos: Results from a family of experiments,” *Information and Software Technology*, vol. 55, no. 10, pp. 1823–1843, 2013.
- [22] Y.-C. Tu, E. Tempero, and C. Thomborson, “An experiment on the impact of transparency on the effectiveness of requirements documents,” *Empirical Software Engineering*, 2015.

-
- [23] A. Ahmad, P. Jamshidi, and C. Pahl, “Protocol for Systematic Literature Review,” 2012.
- [24] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering,” *Engineering*, vol. 2, p. 1051, 2007.
- [25] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, “Requirements engineering paper classification and evaluation criteria: A proposal and a discussion,” *Requirements Engineering*, vol. 11, pp. 102–107, 2006.
- [26] L. Gurses, “10 mistakes in transitioning to agile,” 2006.
- [27] A. M. R. Da Cruz, “A pattern language for use case modeling,” in *MODELSDWARD 2014 - Proceedings of the 2nd International Conference on Model-Driven Engineering and Software Development*, pp. 408–414, SciTePress, 2014.
- [28] W. Hudson, “Adopting User-Centered Design within an Agile Process: A Conversation,” *Cutter IT Journal*, vol. 16, no. 10, pp. 5–12, 2003.
- [29] A. Elssamadisy, *Agile Adoption Patterns: A Roadmap to Organizational Success*. Addison-Wesley Professional, 1 ed., 2008.
- [30] J. Nawrocki, M. Ochodek, J. Jurkiewicz, S. Kopczyńska, and B. Alchimowicz, “Agile requirements engineering: A research perspective,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8327 LNCS, pp. 40–51, 2014.
- [31] V. Laporti, M. R. Borges, and V. Braganholo, “Athena: A collaborative approach to requirements elicitation,” *Computers in Industry*, vol. 60, pp. 367–380, aug 2009.
- [32] L. Antonelli, G. Rossi, J. C. Sampaio do Prado Leite, and A. Oliveros, “Deriving requirements specifications from the application domain language captured by Language Extended Lexicon,” in *WER 2012 - 15th Workshop on Requirements Engineering*, 2012.
- [33] T. Ishikuma, “International standardization for realizing intensive information oriented, safe and secure low-carbon + smart society - System interface between industrial facilities and the smart grid,” in *Proceedings of the SICE Annual Conference*, pp. 1–4, 2013.
- [34] Y. Bugayenko, “Method for software cost estimating using scope champions,” in *Lecture Notes in Business Information Processing* (F. Boma-rius, M. Oivo, P. Jaring, and P. Abrahamsson, eds.), vol. 32 of *Lecture Notes in Business Information Processing*, (Berlin, Heidelberg), pp. 59–70, Springer Berlin Heidelberg, 2009.

Bibliografía

- [35] C. Riedemann and R. Freitag, “Modeling usage: Techniques and tools,” *IEEE Software*, vol. 26, no. 2, pp. 20–24, 2009.
- [36] W. M. Farid and F. J. Mitropoulos, “NORMATIC: A visual tool for modeling Non-Functional Requirements in agile processes,” in *2012 Proceedings of IEEE Southeastcon*, pp. 1–8, IEEE, mar 2012.
- [37] W. M. Farid and F. J. Mitropoulos, “Novel lightweight engineering artifacts for modeling non-functional requirements in agile processes,” in *2012 Proceedings of IEEE Southeastcon*, pp. 1–7, IEEE, mar 2012.
- [38] P. Kamthan, “On the suitability of the social Web environment for agile requirements engineering,” in *Lecture Notes in Electrical Engineering* (J. J. Park, V. C. Leung, C.-L. Wang, and T. Shon, eds.), vol. 179 of *Lecture Notes in Electrical Engineering*, (Dordrecht), pp. 193–200, Springer Netherlands, 2012.
- [39] J. Drobka, D. Noftz, and R. Raghu, “Piloting XP on four mission-critical projects,” *IEEE Software*, vol. 21, no. 6, pp. 70–75, 2004.
- [40] R. E. Gallardo-Valencia, V. Olivera, and S. E. Sim, “Practical experiments are informative, but never perfect,” in *Proceedings of the 1st ACM international workshop on Empirical assessment of software engineering languages and technologies held in conjunction with the 22nd IEEE/ACM International Conference on Automated Software Engineering (ASE) 2007 - WEASEL Tech '07*, (New York, New York, USA), pp. 37–42, ACM Press, 2007.
- [41] J. Archer, “Requirement tracking: a streamlined approach,” *Proceedings. 11th IEEE International Requirements Engineering Conference, 2003.*, p. 1, 2003.
- [42] V. Monochristou and M. Vlachopoulou, “Requirement Specification using User Stories,” in *Agile Software Development Quality Assurance* (I. G. Stamelos and P. Sfetsos, eds.), pp. 71–89, IGI Global, 2007.
- [43] M. Smialek, J. Bojarski, W. Nowakowski, T. Straszak, M. Śmiałek, J. Bojarski, W. Nowakowski, and T. Straszak, “Scenario construction tool based on extended UML metamodel,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3713 LNCS, pp. 414–429, 2005.
- [44] ISO, IEC, and IEEE, “Systems and software engineering – Developing user documentation in an agile environment,” 2012.

-
- [45] W. M. Farid, *The Normap Methodology: Non-functional Requirements Modeling for Agile Processes*. PhD thesis, Nova Southeastern University, 2011.
- [46] I. Hussain, L. Kosseim, and O. Ormandjieva, “Towards approximating COSMIC functional size from user requirements in agile development processes using text mining,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (C. J. Hopfe, Y. Rezgui, E. Métais, A. Preece, and H. Li, eds.), vol. 6177 of *Lecture Notes in Computer Science*, (Berlin, Heidelberg), pp. 80–91, Springer Berlin Heidelberg, 2010.
- [47] M. Kumar, N. Ajmeri, and S. Ghaisas, “Towards knowledge assisted agile requirements evolution,” in *Proceedings of the 2nd International Workshop on Recommendation Systems for Software Engineering - RS-SE '10*, (New York, New York, USA), pp. 16–20, ACM Press, 2010.
- [48] M. Imaz, “Zooming in and out in requirements engineering,” in *Lecture Notes in Business Information Processing* (X. Franch and P. Soffer, eds.), vol. 148 of *Lecture Notes in Business Information Processing*, (Berlin, Heidelberg), pp. 174–185, Springer Berlin Heidelberg, 2013.
- [49] W. M. Farid, “The Normap methodology: Lightweight engineering of non-functional requirements for agile processes,” in *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, vol. 1, pp. 322–325, 2012.
- [50] O. Liskin, “How Artifacts Support and Impede Requirements Communication,” in *21st International Working Conference, REFSQ 2015*, vol. 9013, pp. 132–147, 2015.
- [51] W. M. Farid and F. J. Mitropoulos, “Visualization and scheduling of non-functional requirements for agile processes,” in *Conference Proceedings - IEEE SOUTHEASTCON*, 2013.
- [52] M. Zarour, A. Abran, J.-M. Desharnais, and A. Alarifi, “An investigation into the best practices for the successful design and implementation of lightweight software process assessment methods: A systematic literature review,” *Journal of Systems and Software*, vol. 101, pp. 180–192, 2015.
- [53] A. Jedlitschka, M. Ciolkowski, and D. Pfahl, “Reporting experiments in software engineering,” in *Guide to Advanced Empirical Software Engineering*, pp. 201–228, 2008.
- [54] F. D. Davis, “Perceived Ease of Use, and User Acceptance of Information Technology,” *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

Bibliografía

- [55] B. Kitchenham, S. Pfleeger, L. Pickard, P. Jones, D. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering," *IEEE Transactions on Software Engineering*, vol. 28, no. 8, pp. 721–734, 2002.
- [56] R. Likert, "A technique for the measurement of attitudes.," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.
- [57] I. E. Allen and C. A. Seaman, "Likert scales and data analyses," 2007.
- [58] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.
- [59] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.
- [60] Student, "The Probable Error of a Mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.
- [61] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [62] F. Wilcoxon, "Individual comparisons by ranking methods.," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.
- [63] D. G. Bonett and T. A. Wright, "Sample size requirements for estimating Pearson, Kendall and Spearman correlations," *Psychometrika*, vol. 65, no. 1, pp. 23–28, 2000.
- [64] H. E. Soper, A. W. Young, B. M. Cave, A. Lee, and K. Pearson, "On the Distribution of the Correlation Coefficient in Small Samples. Appendix II to the Papers of "Student."and R. A. Fisher," *Biometrika*, vol. 11, no. 4, pp. 328–413, 1917.
- [65] G. U. Yule, "Notes of Karl Pearson's Lectures on the Theory of Statistics," *Biometrika*, vol. 30, no. 1, pp. 198–203, 1938.
- [66] C. Spearman, "The Proof and Measurement of Association between Two Things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
- [67] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix.," *Psychological bulletin*, vol. 56, no. 2, pp. 81–105, 1959.

Bibliografía

- [68] D. W. Straub, “Validating Instruments in MIS Research,” *MIS Quarterly*, vol. 13, no. 2, pp. 147–169, 1989.
- [69] J. C. Nunnally, “Psychometric Theory,” *Educational Researcher*, vol. 4, no. 10, pp. 7–14+19–21, 1975.
- [70] D. Muijs, *Doing quantitative research in education with SPSS*. Sage, second edi ed., 2010.



Bibliografía

